

# Fast Adaptive S-ALOHA Scheme for Event-driven M2M Communications

Huasen Wu<sup>\*†</sup>, Chenxi Zhu<sup>‡</sup>, Richard J. La<sup>§</sup>, Xin Liu<sup>†</sup>, and Youguang Zhang<sup>\*</sup>

<sup>\*</sup>School of Electronic and Information Engineering, Beihang University, Beijing

<sup>†</sup>Department of Computer Science, University of California, Davis

<sup>‡</sup>Mallard Creek Networks, 11452 Mallard Creek Trail, Fairfax, VA

<sup>§</sup>Department of Electrical and Computer Engineering, University of Maryland, College Park, MD

**Abstract**—Supporting massive device transmission is challenging in Machine-to-Machine (M2M) communications. Particularly, in event-driven M2M communications, a large number of devices activate within a short period of time, which in turn causes high radio congestions and severe access delay. To address this issue, we propose a Fast Adaptive S-ALOHA (FASA) scheme for random access control of M2M communication systems with bursty traffic. Instead of the observation in a single slot, the statistics of consecutive idle and collision slots are used in FASA to accelerate the tracking process of network status which is critical for optimizing S-ALOHA systems. Using drift analysis, we design the FASA scheme such that the estimate of the backlogged devices converges fast to the true value. Furthermore, by examining the  $T$ -slot drifts, we prove that the proposed FASA scheme is stable as long as the average arrival rate is smaller than  $e^{-1}$ , in the sense that the Markov chain derived from the scheme is geometrically ergodic. Simulation results demonstrate that the proposed FASA scheme outperforms traditional additive schemes such as PB-ALOHA and achieves near-optimal performance in reducing access delay. Moreover, compared to multiplicative schemes, FASA shows its robustness under heavy traffic load in addition to better delay performance.

**Index Terms**—M2M communications, random access control, adaptive S-ALOHA, drift analysis, stability analysis.

## I. INTRODUCTION

**M**ACHINE-to-Machine (M2M) communication or Machine-Type Communication (MTC) is expected to be one of the major drivers of cellular networks [1] and has become one of the focuses in 3GPP [2–4]. Behind the proliferation of M2M communication, the congestion problems in M2M communication become a big concern. The reason is that the device density of M2M communication is much higher than that in traditional Human-to-Human (H2H) communication [2, 3]. For example, it is expected in [3] that 1,000 devices/km<sup>2</sup> are deployed for environment monitoring and control. What is worse, in event-driven M2M applications, many devices may be triggered almost simultaneously and attempt to access the base station (BS) through the Random Access Channel (RACH) [5]. Such high burstiness can result in congestion and increase access delay, which motivates our research.

Several efforts have been made in 3GPP to alleviate the radio congestion on RACH. Back to 3GPP LTE (Long Term Evolution), since the amount of exchanged data grows rapidly, congestion control has been addressed and access class barring (ACB) scheme has been proposed for overload protection [6]. In ACB scheme, devices are divided into several access classes. Before establishing a connection, the device is required to perform the ACB check and randomly transmit

request packets with a probability broadcasted by the BS. Therefore, the traffic load can be reduced by choosing a small transmission probability. It is noticed that ACB scheme is a good candidate for congestion control in M2M applications, though modifications in accordance with the specific features of M2M applications is necessary [7]. Therefore, in [8], a two stage access control scheme, which consists of a primary level and a secondary level of access control barring, is introduced to provide prioritized M2M services based on their service attributes. Moreover, the authors in [9] propose a cooperative ACB scheme for balancing traffic load among BSs in a heterogeneous multi-tier cellular network. With cooperations among BSs, the congestion level can be reduced and the access delay can be significantly improved. However, a key problem arising in implementing these schemes is how to estimate the number of active devices and optimize the transmission probability. This problem becomes even worse in event-driven M2M applications which is characterized with highly bursty traffic.

Essentially, the ACB scheme belongs to slotted-ALOHA (S-ALOHA) type schemes, which are widely applied for random access control. To address the instability issue of S-ALOHA [10], plenty of work has been done for deciding the protocol parameters and stabilizing the S-ALOHA system, which is briefly summarized in Section II. In these schemes, historical outcomes are applied to estimate the network status and adjust protocol parameters. Drift analysis is then used to design the schemes and prove their stability [11–13]. However, these schemes usually rely on the assumption that the traffic can be modeled as a Poisson process and only apply the observation in the previous slot for estimating the number of backlogged devices. Due to the burstiness, this assumption cannot be justified in the context of M2M applications and the observation in a single slot is not satisfactory for adjusting the protocol parameters in time. Thus, we try to make full use of the information provided by the historical outcomes for improving the performance of access control under bursty traffic.

In this paper, we study adaptive S-ALOHA scheme for event-driven M2M communication and provide rigorous analysis about its stability. As our main contribution, we propose a Fast Adaptive S-ALOHA (FASA) scheme for the random access control of M2M devices. A key characteristic of FASA is that the access results in the past slots, in particular, consecutive idles or collisions, are collected and applied to estimate the number of backlogged devices. This enables the fast update of transmission probability under highly bursty traffic and thus reduce the access delay. Furthermore, we prove

the stability of FASA under bursty traffic. This is accomplished by examining the  $T$ -slot drifts of the network status, which captures the memory property of FASA. Under interrupted Poisson traffic model [14], we show that the  $T$ -slot drifts of FASA have the required properties for stabilizing the scheme and the system is stable when the arrival rate is less than  $e^{-1}$ . Numerical results demonstrate that using FASA scheme, the transmission probability of S-ALOHA can be adjusted in time and the access delay can be reduced to be very close to the theoretical lower bound under highly bursty traffic.

The remainder of the paper is organized as follows. Section II summarizes the related work. In Section III, we present the system model, including the bursty traffic model for the event-driven M2M communications. In Section IV, after analyzing the limitations of traditional fixed step size policies, we propose the FASA scheme and design the parameters in the scheme based on drift analysis. Then in Section V, we study the  $T$ -slot drifts of FASA and prove its stability. In Section VI, simulation results are presented to evaluate the performance of the proposed scheme, compared with the theoretical optimal scheme in ideal case and two traditional adaptive schemes. Finally, we conclude our paper in Section VII.

## II. RELATED WORK

*Radio Congestion Control in M2M communications:* Due to the high equipment density, the radio congestion of M2M applications is still an open problem. Besides the efforts in 3GPP [7, 9, 15], there are a few publications addressing this issue. Since group-based feature appears in many M2M applications, some researchers propose hierarchical architectures, such as grouping scheme [16] and relay schemes [17, 18], for alleviating the radio congestion on the RACH. In these architectures, group heads are selected for collecting the messages from the group members. However, efficient schemes for communications between the group head and members is required to reduce the access delay. At the same time, Adaptive Traffic Load Slotted Multiple Access Collision Avoidance (ATL S-MACA) mechanism proposed in [19] uses packet sensing and adaptive method to improve the access performance under high traffic load. However, the scheme is designed for Poisson traffic and is not suitable for event-driven M2M applications.

*Adaptive S-ALOHA Schemes:* S-ALOHA is a fundamental scheme for random access control and the combinations with other techniques such as CDMA make it even more useful. However, the instability issue of S-ALOHA should be dealt with when being implemented in real networks. Two typical classes of schemes, additive and multiplicative adaptive schemes, have been proposed for stabilizing S-ALOHA systems [20]. In these schemes, the estimate of the network status is updated in an additive or multiplicative manner, respectively, and the transmission probability is adjusted accordingly. However, as discussed in more detail later, traditional additive schemes such as Pseudo Bayesian ALOHA (PB-ALOHA) [21] estimate the number of backlogged devices based on the access result in the previous slot but cannot adjust the transmission probability in time under highly bursty traffic, resulting in large access delay. On the other hand, because of the exponential increment in consecutive collision slots or decrement in consecutive idle slots, multiplicative schemes [22], e.g., Q-Algorithm in [23] and its enhanced version  $Q^+$ -Algorithm in [24], can track the network status in a short

period. However, the throughput suffers in these schemes due to the fluctuations in the estimation [22]. Therefore, we aim to design adaptive schemes that could adjust the protocol parameters fast under bursty traffic while retaining the same stable throughput as additive schemes. In our previous work [25], we propose a preliminary version of FASA based on some intuitive approximations and show its desirable properties through numerical simulations. However, no rigorous analysis about the stability of FASA is presented in [25].

*Drift Analysis for stabilization of S-ALOHA:* Drift analysis is a theory for deducing the properties such as ergodicity of a sequence from its drift and is found useful in the design and analysis of adaptive S-ALOHA schemes [12, 26, 27]. The network status, which is represented by the number of backlogged devices and its estimate, could be viewed as a stochastic sequence. It is shown in [12] that when the drifts of the network status satisfies some criterions, the system is stable in the sense that the returning time can be bounded with high probability. Using the conclusion in [12], the most related work [13] studies the stability of PB-ALOHA scheme by defining a Lyapunov function to represent the network status and examining its drift. In all the work mentioned above, the schemes update the parameters based on the observation in the previous slot. Thus, the 1-slot drifts, i.e., drifts between two adjacent slots, are sufficient for studying the stability of the systems. When involving access results in multiple slots, however, the  $T$ -slot drifts are required to deal with the memory property of our scheme. Unlike 1-slot drifts, calculating  $T$ -slot drifts is non-trivial and we have to resort to approximations for obtaining their properties.

## III. SYSTEM MODEL

In this section, we describe random access control procedure for M2M communication as well as the traffic model, which will be used for studying the stability and evaluating the performance of the proposed scheme.

### A. S-ALOHA Based Random Access

We consider a cellular network based M2M communication system for event detection. As shown in Fig. 1, the system consists of a BS and a large number of M2M devices. When an event is detected, certain number of devices are triggered and attempt to access the BS by sending request packets through a single RACH based on S-ALOHA scheme.

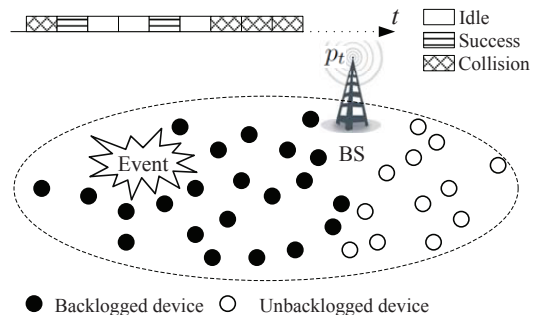


Fig. 1. S-ALOHA based access control of M2M communications

The time is divided into time slots, each of which is long enough to transmit a request packet. Deferred first transmission (DFT) mode [21] is assumed, in which a device with a new

request packet immediately goes to backlogged state. In the  $t$ th slot, where  $t \in \mathbb{Z}_+ := \{0, 1, 2, \dots\}$ , all backlogged devices transmit packets with probability  $p_t$ , which is broadcasted by the BS at the beginning of the slot. For the sake of simplicity, we assume that the request packets generated by M2M devices will eventually be transmitted successfully and a backlogged device will not generate any new requests since the new coming data can be transmitted as long as the device accesses the BS successfully. Moreover, an ideal collision channel is assumed, where the transmitted packet will be successfully received by the BS when no other packets are being transmitted in the same slot.

The transmission probability  $p_t$  is adjusted based on access results in the past. Let  $Z_t$  denote the access result in slot  $t$ , and  $Z_t = 0, 1$ , or  $c$  depending on whether zero, one, or more than one request packets are transmitted on the RACH. At the end of slot  $t$ , the BS decides the transmission probability for next slot based on the sequence  $\{Z_0, Z_1, \dots, Z_t\}$ , i.e.,

$$p_{t+1} = \Gamma_t(Z_0, Z_1, \dots, Z_t).$$

The objective of the BS is to maximize the throughput and minimize the access delay. It well known that, when  $N_t \geq 1$ , where  $N_t$  is the number of backlogged devices in slot  $t$ , using a transmission probability  $p_t = 1/N_t$  in slot  $t$  maximizes the throughput of the S-ALOHA system. However, the BS does not know  $N_t$  and has to obtain its estimate  $\hat{N}_t$  based on the access results in the past.

### B. Traffic Model

In order to capture the burstiness of event-driven M2M traffic, instead of traditional Poisson process, the arrival process is modeled as an *interrupted* Poisson process, which was suggested by Hayward of Bell Laboratories for simulating overflow traffic [14].

Interrupted Poisson process can be viewed as a Poisson process modulated by a random switch and will be discretized according to the slotted structure of the scheme. Let  $Y_t$  and  $A_t$  respectively denote the number of events happening and the number of devices triggered in slot  $t$ . Assume that events happen independently and identically in each slot and at most one event happens in one slot, with  $\theta$  being the happening probability. Hence, in each slot  $t$ ,  $\Pr(Y_t = 1) = \theta$  and  $\Pr(Y_t = 0) = 1 - \theta$ . In addition, assume that the number of triggered devices follows Poisson distribution with mean  $\lambda$  when an event happens, and no devices become active otherwise, i.e.,  $A_t \sim \mathcal{P}(\lambda)$  when  $Y_t = 1$  (ON-state), and  $A_t = 0$  when  $Y_t = 0$  (OFF-state). Therefore, random variables  $\{A_t : t \in \mathbb{Z}_+\}$  are independently and identically distributed (i.i.d.) and the long term arrival rate can be calculated as

$$\bar{\lambda} = \Pr(Y_t = 0) \cdot 0 + \Pr(Y_t = 1) \cdot \lambda = \theta\lambda. \quad (1)$$

In addition, the burstiness of the traffic is reflected by the variance of  $A_t$ , which is given by

$$\sigma_A^2 = \theta(\lambda + \lambda^2) - (\theta\lambda)^2 = \bar{\lambda}(1 + \lambda - \bar{\lambda}). \quad (2)$$

We note that the classic Poisson process is included as a special case of this model when  $\theta = 1$ . We will design and analyze adaptive S-ALOHA scheme based on this traffic model. Indeed, as will be discussed later, the scheme proposed in this paper could be stable under some other more general traffic models.

## IV. FAST ADAPTIVE S-ALOHA

The estimation of the number of backlogged devices plays an important part in stabilizing and optimizing the S-ALOHA system. In this section, using drift analysis, we first examine the limit of traditional fixed step size estimation schemes. Then, we propose and analyze a fast adaptive scheme, referred to as Fast Adaptive S-ALOHA.

### A. Drift Analysis of Fixed Step size Estimation Schemes

Many schemes with fixed step size have been proposed in the literature to estimate the number of backlogged devices [20]. A unified framework of additive schemes is proposed and studied by Kelly in [11], where the estimate  $\hat{N}_t$  is updated by the recursion

$$\hat{N}_{t+1} = \max\{1, \hat{N}_t + a_0 I(Z_t = 0) + a_1 I(Z_t = 1) + a_c I(Z_t = c)\}, \quad (3)$$

where  $a_0$ ,  $a_1$ , and  $a_c$  are constants and  $I(A)$  is the indicator function of event  $A$ .

With the estimation, the BS sets the transmission probability to  $p_t = 1/\hat{N}_t$  for all backlogged devices, and thus the offered load  $\rho = N_t p_t = N_t / \hat{N}_t$ , representing the average number of devices attempting to access the channel. To stabilize and optimize the S-ALOHA system,  $\hat{N}_t$  needs to drift towards the actual number of backlogged devices  $N_t$ , especially when  $N_t$  is large. When  $N_t = n$  and  $\hat{N}_t = \hat{n}$ , the drift of the estimate can be calculated as follows [11]:

$$\begin{aligned} & E[\hat{N}_{t+1} - \hat{N}_t | N_t = n, \hat{N}_t = \hat{n}] \\ &= (a_0 - a_c) \left(1 - \frac{1}{\hat{n}}\right)^n + (a_1 - a_c) \frac{n}{\hat{n}} \left(1 - \frac{1}{\hat{n}}\right)^{n-1} + a_c \\ &\rightarrow (a_0 - a_c)e^{-\rho} + (a_1 - a_c)\rho e^{-\rho} + a_c \stackrel{\text{def}}{=} \phi(\rho), \end{aligned}$$

as  $n \rightarrow \infty$ , with  $n/\hat{n} = \rho$  fixed.

By properly choosing the parameters  $a_i$  ( $i = 0, 1, c$ ) such that  $\phi(\rho) < 0$  if  $\rho < 1$  and  $\phi(\rho) > 0$  if  $\rho > 1$ , the estimate  $\hat{N}_t$  will drift towards the true value and thus the S-ALOHA system can be stabilized. However, these fixed step size schemes are not suitable for systems with bursty traffic. When the estimate  $\hat{N}_t$  deviates far away from the true value  $N_t$ , we have  $\lim_{\rho \rightarrow 0} \phi(\rho) = a_0$  and  $\lim_{\rho \rightarrow \infty} \phi(\rho) = a_c$ . These limits indicate that the drift tends to be a constant even when the deviation is large, which could result in a large tracking time. Thus, it is necessary to design fast estimation schemes for event-driven M2M communication.

### B. Framework of FASA

As analyzed in the previous subsection, fixed step size estimation schemes such as PB-ALOHA may not be able to adapt in a timely manner for systems with bursty traffic because it always uses a constant step size even when the estimate is far away from the true value. We note that in addition to the access result in the previous slot, the access results in several consecutive slots will be helpful for improving the estimation as they could reveal additional information about the true value. Intuitively, collisions in several consecutive slots are likely caused by a significant underestimation, i.e.,  $\hat{N}_t \ll N_t$ , and the BS should aggressively increase its estimate. In contrast, several consecutive idle slots may indicate that the estimate  $\hat{N}_t \gg N_t$ , and it should be reduced aggressively.



Motivated by this intuition, we propose a FASA scheme that updates  $\hat{N}_t$  as follows:

$$\hat{N}_{t+1} = \begin{cases} \max\{1, \hat{N}_t - 1 - h_0(\nu)(K_{0,t} \wedge k_m)^\nu\}, & \text{if } Z_t = 0 \\ \hat{N}_t, & \text{if } Z_t = 1 \\ \hat{N}_t + \frac{1}{e-2} + h_c(\nu)(K_{c,t} \wedge k_m)^\nu, & \text{if } Z_t = c \end{cases} \quad (4)$$

where  $K_{0,t}$  and  $K_{c,t}$  are the numbers of consecutive idle and collision slots up to slot  $t$ , respectively;  $k_m > 1$  is an integer and  $K \wedge k_m = \min\{K, k_m\}$ ;  $\nu > 0$  is the parameter that controls the adjusting speed;  $h_0(\nu)$  and  $h_c(\nu)$  are functions of  $\nu$  that guarantee the right direction of the estimation drift. In order to make the scheme implementable and its stability analysis tractable, we bound the update step size with  $k_m$  in this paper, which is different from that we proposed in [25]. However, the two schemes are almost the same as long as we choose a sufficiently large  $k_m$ .

### C. Design of $h_0(\nu)$ and $h_c(\nu)$

The functions  $h_0(\nu)$  and  $h_c(\nu)$  are crucial for guaranteeing the convergence of the FASA scheme. Next we design  $h_0(\nu)$  and  $h_c(\nu)$  by analyzing the drift of estimate  $\hat{N}_t$ . According to the structure of the proposed scheme, the evolution of  $\hat{N}_t$  depends not only on the access result in the previous slot, but also results in the past  $k_m$  slots. Therefore, unlike the fixed step size schemes, accurate drift analysis is impractical for FASA because of its memory property. Thus, in order to make the problem tractable, we resort to approximation based on Lemma 1, which indicates the feasibility for approximating the distribution of access results in the past  $k_m$  slots with the network status in the previous slot.

**Lemma 1** For given  $\epsilon > 0$ ,  $\Delta n$ , and  $\Delta \hat{n}$ , there exists some  $M > 0$ , such that for any  $(n, \hat{n}) \in H_M$ , where  $H_M = \{(n, \hat{n}) : n \geq M \text{ or } \hat{n} \geq M, n + \Delta n \geq 0, \hat{n} + \Delta \hat{n} \geq 1\}$ , we have

$$\left| e^{-\rho} - \left(1 - \frac{1}{\hat{n} + \Delta \hat{n}}\right)^{n + \Delta n} \right| \leq \epsilon, \quad (5)$$

$$\left| \rho e^{-\rho} - \frac{n + \Delta n}{\hat{n} + \Delta \hat{n}} \left(1 - \frac{1}{\hat{n} + \Delta \hat{n}}\right)^{n + \Delta n - 1} \right| \leq \epsilon, \quad (6)$$

where  $\rho = n/\hat{n}$ .

*Proof:* See Appendix A.

It is noticed that the distribution of the access result  $Z_s$  in slot  $s$  is decided by  $N_s$  and  $\hat{N}_s$ . In addition, for any given  $s \in \{t - k_m, t - k_m + 1, \dots, t - 1\}$ , we have  $|\hat{N}_s - \hat{N}_t| \leq k_m/(e - 2) + k_m^{1+\nu} \max\{h_0(\nu), h_c(\nu)\}$ , and  $|N_s - N_t|$  is bounded with high probability, i.e.,  $\Pr(|N_s - N_t| \leq B) \rightarrow 1$  as  $B \rightarrow \infty$ . Thus, according to Lemma 1, when  $N_t$  and  $\hat{N}_t$  are known and at least one of them is sufficiently large, the distribution of access results in the past  $k_m$  slots can be evaluated approximately, as well as the statistical characteristics of  $K_{0,t}$  and  $K_{c,t}$ . Therefore, in this section, we do not assume any knowledge about  $K_{0,t}$  or  $K_{c,t}$  in slot  $t$  and will approximately calculate the drift of  $\hat{N}_t$  conditioned on  $N_t$  and  $\hat{N}_t$ . In addition, we assume that  $\hat{N}_t$  is large enough in the past  $k_m$  slots, and hence we can approximate  $\max\{1, x\}$  in (4) as  $x$  in the analysis later. Rigorous analysis provided in Section V will show that the design with these approximations stabilizes the proposed scheme.

Suppose that in slot  $t$ , the number of backlogged devices and its estimate are  $N_t = n$ ,  $\hat{N}_t = \hat{n}$ , respectively, and thus the offered load  $\rho = n/\hat{n}$ . When  $n$  or  $\hat{n}$  is large, the drift of estimate  $\hat{N}_t$  can be approximated as

$$\begin{aligned} E[\hat{N}_{t+1} - \hat{N}_t | N_t = n, \hat{N}_t = \hat{n}] \\ \approx \sum_{i \in \{0,1,c\}} q_i(\rho) E[\Delta \hat{N}_t | (i, n, \hat{n})] \end{aligned} \quad (7)$$

where  $q_0(\rho) = e^{-\rho}$ ,  $q_1(\rho) = \rho e^{-\rho}$ , and  $q_c(\rho) = 1 - q_0(\rho) - q_1(\rho)$  are the probabilities of an *idle*, *success*, and *collision* slot, respectively;  $E[\Delta \hat{N}_t | (i, n, \hat{n})]$  ( $i = 0, 1, c$ ) are the changes in  $\hat{N}_t$  resulting from the corresponding updates.

Obviously,  $E[\Delta \hat{N}_t | (1, n, \hat{n})] = 0$  since the estimated number remains unchanged when a packet is successfully transmitted in slot  $t$ . On the other hand, without memory about the access results in the past slots,  $K_{0,t}$  and  $K_{c,t}$  are treated as random variables. Hence,  $E[\Delta \hat{N}_t | (0, n, \hat{n})]$  and  $E[\Delta \hat{N}_t | (c, n, \hat{n})]$  can be obtained based on the approximate distributions of  $K_{0,t}$  and  $K_{c,t}$ .

First, to calculate the drift of estimate in an idle slot  $E[\Delta \hat{N}_t | (0, n, \hat{n})]$ , suppose that no packet is transmitted in slot  $t$ , then the estimated number will be reduced by  $1 + h_0(\nu)(K_{0,t} \wedge k_m)^\nu$ . Therefore,

$$\begin{aligned} E[\Delta \hat{N}_t | (0, n, \hat{n})] \\ = - \sum_{k_0=1}^{k_m-1} [1 + h_0(\nu)k_0^\nu] \Pr[K_{0,t} = k_0 | (0, n, \hat{n})] \\ - [1 + h_0(\nu)k_m^\nu] \Pr[K_{0,t} \geq k_m | (0, n, \hat{n})]. \end{aligned} \quad (8)$$

Notice that  $K_{0,t} = k_0$  ( $1 \leq k_0 < k_m$ ) holds when slots  $t - k_0 + 1, t - k_0 + 2, \dots, t - 1$  are all idle while slot  $t - k_0$  is not. Thus, for  $1 \leq k_0 < k_m$ , we have

$$\begin{aligned} \Pr[K_{0,t} = k_0 | (0, n, \hat{n})] \\ = \Pr[Z_{t-k_0} \neq 0 | (Z_{t-k_0+1}, \dots, Z_t, N_t, N'_t) = (0, \dots, 0, n, \hat{n})] \\ \cdot \prod_{s=t-k_0+1}^{t-1} \Pr[Z_s = 0 | (Z_{s+1}, \dots, Z_t, N_t, N'_t) = (0, \dots, 0, n, \hat{n})]. \end{aligned}$$

According to Lemma 1, when  $N_t = n$  or  $\hat{N}_t = \hat{n}$  are sufficiently large, the distribution of access results in the past  $k_m$  slots can be approximated as that in the previous slot, i.e.,

$$\begin{aligned} \Pr[Z_{t-k_0} \neq 0 | (Z_{t-k_0+1}, \dots, Z_t, N_t, N'_t) = (0, \dots, 0, n, \hat{n})] \\ \approx 1 - q_0(\rho), \\ \Pr[Z_s = 0 | (Z_{s+1}, \dots, Z_t, N_t, N'_t) = (0, \dots, 0, n, \hat{n})] \\ \approx q_0(\rho), \quad s = t - k_0 + 1, t - k_0 + 2, \dots, t - 1. \end{aligned}$$

Consequently, for  $1 \leq k_0 < k_m$ ,

$$\Pr[K_{0,t} = k_0 | (0, n, \hat{n})] \approx q_0^{k_0-1}(\rho) [1 - q_0(\rho)]. \quad (9)$$

Similarly,  $K_{0,t} \geq k_m$  holds if slots  $t - k_m + 1, t - k_m + 2, \dots, t - 1$  are all idle and we can approximate the probability as

$$\Pr(K_{0,t} \geq k_m) \approx q_0^{k_m-1}(\rho). \quad (10)$$

Substituting (9) and (10) into (8), we can calculate the drift of  $\hat{N}_t$  in an idle slot as follows:

$$\begin{aligned} E[\Delta \hat{N}_t | (0, n, \hat{n})] &\approx - \sum_{k_0=1}^{k_m-1} [1 + h_0(\nu)k_0^\nu] q_0^{k_0-1}(\rho) [1 - q_0(\rho)] \\ &\quad - [1 + h_0(\nu)k_m^\nu] q_0^{k_m-1}(\rho) \\ &= -[1 + h_0(\nu)\mu(\nu, q_0(\rho), k_m)], \end{aligned} \quad (11)$$

where  $\mu(\nu, q, k_m)$  is defined as

$$\mu(\nu, q, k_m) = \sum_{k=1}^{k_m-1} k^\nu q^{k-1} (1-q) + (k_m)^\nu q^{k_m-1}, \quad (12)$$

and  $\mu(\nu, q_0(\rho), k_m)$  is the approximate expectation of  $(K_{0,t} \wedge k_m)^\nu$  conditioned on  $(Z_t, N_t, \hat{N}_t) = (0, n, \hat{n})$ .

Second, we can calculate the drift of estimate in a collision slot in a similar fashion as follows:

$$\begin{aligned} E[\Delta \hat{N}_t | (c, n, \hat{n})] &\approx (e-2)^{-1} + h_c(\nu) E[(K_{c,t} \wedge k_m)^\nu | (c, n, \hat{n})] \\ &= (e-2)^{-1} + h_c(\nu) \mu(\nu, q_c(\rho), k_m). \end{aligned} \quad (13)$$

Therefore, the drift of estimate for FASA can be approximated by substituting the expressions of  $E[\Delta \hat{N}_t | (i, n, \hat{n})]$  ( $i = 0, 1, c$ ) into (7):

$$\begin{aligned} E[\hat{N}_{t+1} - \hat{N}_t | N_t = n, \hat{N}_t = \hat{n}] &\approx -q_0(\rho) [1 + h_0(\nu)\mu(\nu, q_0(\rho), k_m)] \\ &\quad + q_c(\rho) [(e-2)^{-1} + h_c(\nu)\mu(\nu, q_c(\rho), k_m)] \\ &\stackrel{\text{def}}{=} \varphi(\rho). \end{aligned} \quad (14)$$

In order to keep the offered load  $\rho$  staying in the neighborhood of the optimal value  $\rho^* = 1$ , it is reasonable to require that  $\varphi(1) = 0$ . In other words, letting  $q_0^* = q_0(1) = e^{-1}$  and  $q_c^* = q_c(1) = 1 - 2e^{-1}$ , we expect that

$$\begin{aligned} \varphi(1) &= -q_0^* [1 + h_0(\nu)\mu(\nu, q_0^*, k_m)] \\ &\quad + q_c^* [(e-2)^{-1} + h_c(\nu)\mu(\nu, q_c^*, k_m)] \\ &= -h_0(\nu)q_0^*\mu(\nu, q_0^*, k_m) + h_c(\nu)q_c^*\mu(\nu, q_c^*, k_m) = 0. \end{aligned} \quad (15)$$

Hence, for given  $k_m > 1$ , to satisfy the condition in (15), we can select the following  $h_0(\nu)$  and  $h_c(\nu)$ :

$$h_0(\nu) = \eta [q_0^*\mu(\nu, q_0^*, k_m)]^{-1}, \quad (16)$$

$$h_c(\nu) = \eta [q_c^*\mu(\nu, q_c^*, k_m)]^{-1}, \quad (17)$$

where  $\eta > 0$  is a constant.

The chosen  $h_0(\nu)$  and  $h_c(\nu)$  guarantee that  $\varphi(1) = 0$  and thus provide a necessary condition for FASA to track the number of backlogged devices. Furthermore, Theorem 1 shows a desirable property of FASA, with which the estimated number  $\hat{N}_t$  roughly drifts towards to the true value  $N_t$ , eventually yielding  $\rho = N_t / \hat{N}_t \approx 1$ .

**Theorem 1** *Given that  $h_0(\nu)$  and  $h_c(\nu)$  are defined by (16) and (17), respectively, the approximate drift of FASA  $\varphi(\rho)$  is a strictly increasing function of  $\rho$ . In addition,  $\varphi(\rho) < 0$  when  $0 < \rho < 1$  and  $\varphi(\rho) > 0$  when  $\rho > 1$ .*

*Proof:* See Appendix B.

In order to understand better the behavior of the scheme, we now present the approximate drift of estimate for FASA

with  $\eta = 1$  and  $\nu = 1, 2, 3$ . Assume that  $k_m$  is large, and thus the distribution of  $K_{i,t}$  ( $i = 0, c$ ) can be approximated as a geometrical distribution with success probability  $1 - q_i(\rho)$ . In addition, for  $\nu \in \mathbb{Z}_+$ ,  $\mu(\nu, q_i(\rho), k_m)$  is approximately the  $\nu$ -th-moment of a geometrically distributed random variable with success probability  $1 - q_i(\rho)$ , and its closed-form expression can be obtained. Consequently, the results obtained in our previous work [25] can be applied directly. Fig. 2 shows the approximate drift of estimation versus offered load, where the subscript of FASA represents value of  $(\eta, \nu)$ . It can be observed from the figure that when the estimated number deviates far away from the actual number of backlogged devices, i.e.,  $\rho \approx 0$  or  $\rho \gg 1$ , FASA adjusts its step size accordingly, while PB-ALOHA still uses the same step size. Therefore, using FASA results in much shorter adjusting time than PB-ALOHA, and thus could improve the performance of M2M communication systems with bursty traffic. Note that the drifts of multiplicative schemes such as  $Q^+$ -Algorithm are not illustrated here since they depend on not only the offered load  $\rho$  but also the estimate  $\hat{N}_t$ .

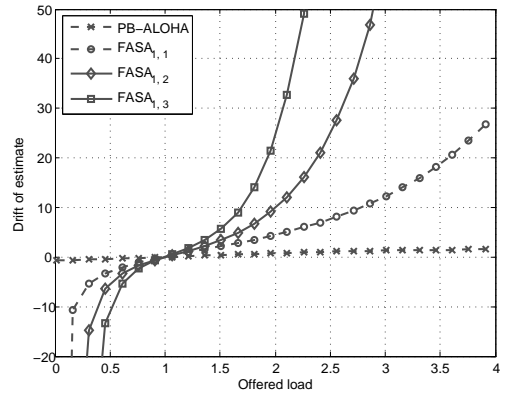


Fig. 2. Drift of estimation.

## V. STABILITY ANALYSIS OF FASA

In this section, we use drift analysis to study the stability of the proposed FASA scheme. The M2M traffic is modeled as an interrupted Poisson process presented in Section III. In fact, as we will discuss later, the proposed scheme can be stable under other more general arrival processes.

Unlike traditional adaptive schemes, the access results in the past consecutive slots are used in FASA to accelerate the speed of tracking, which makes it difficult to obtain the accurate drift of estimate. However, from the stability point of view, we concern mostly the scenarios where the number of backlogged devices or its estimate is large and hence approximation can be applied in these cases. To deal with the issues caused by the memory property of FASA, we analyze its  $T$ -slot drifts rather than the 1-slot drifts, which are introduced for analyzing traditional adaptive ALOHA schemes [12, 13, 22, 27]. By constructing a virtual sequence, we show that the  $T$ -slot drifts of FASA have the properties required for stabilizing the system as long as the number of backlogged or its estimate is sufficiently large, and these are similar to the properties of PB-ALOHA. Therefore, with slight modification, the Lyapunov function based method proposed for PB-ALOHA [13] can be used to prove the stability of FASA.

Consider the FASA scheme proposed in (4) under interrupted Poisson arrival process with average arrival rate  $\bar{\lambda}$ . We define a sequence  $X_t = (N_t, \hat{N}_t, K_t)$ , where  $K_t$  represents the memory of access results in the past consecutive slots and is defined as  $K_0 = 0$  and for  $t > 0$ ,

$$K_t = \begin{cases} -(K_{0,t-1} \wedge k_m), & \text{if } Z_{t-1} = 0, \\ 0, & \text{if } Z_{t-1} = 1, \\ K_{c,t-1} \wedge k_m, & \text{if } Z_{t-1} = c. \end{cases}$$

Recall that  $K_{0,t-1}$  and  $K_{c,t-1}$  is the number of consecutive idle and collision slots up to slot  $t-1$ . Given initial value  $X_0 = (0, 1, 0)$ , each component of  $X_t$  evolves as follows when  $t > 0$ :

$$K_{t+1} = \begin{cases} -(|K_t| + 1) \wedge k_m, & \text{if } K_t < 0, Z_t = 0, \\ -1, & \text{if } K_t \geq 0, Z_t = 0, \\ 0, & \text{if } Z_t = 1, \\ 1, & \text{if } K_t \leq 0, Z_t = c, \\ (K_t + 1) \wedge k_m, & \text{if } K_t > 0, Z_t = c, \end{cases} \quad (18)$$

$$N_{t+1} = \max\{0, N_t + A_t - I(Z_t = 1)\} \quad (19)$$

$$\hat{N}_{t+1} = \begin{cases} \max\{1, \hat{N}_{t+1} - 1 - h_0(\nu)|K_{t+1}|^\nu\}, & \text{if } Z_t = 0, \\ \hat{N}_t, & \text{if } Z_t = 1, \\ \hat{N}_t + \frac{1}{e-2} + h_c(\nu)(K_{t+1})^\nu, & \text{if } Z_t = c. \end{cases} \quad (20)$$

It is easy to verify that  $X_t = (N_t, \hat{N}_t, K_t)$  is a Markov chain on a countable state space  $\mathbb{S}_X$ . The main result of this section reveals the geometrical ergodicity [12] of  $X_t$ , which is described in Theorem 2. As pointed out in [12], the geometrical ergodicity is a weaker form of ergodicity and indicates the existence of steady distribution for each initial state.

**Theorem 2** *If  $0 < \bar{\lambda} < e^{-1}$ ,  $k_m > 1$ ,  $h_0(\nu)$  and  $h_c(\nu)$  are given by (16) and (17), then the Markov Chain  $X_t$  is geometrically ergodic.*

*Proof:* The proof of Theorem 2 is based on the drift analysis. Specifically, the proof involves three steps, which are outlined as follows and presented afterwards:

*Step 1 - Approximation of drifts:* To deal with the impact of the memory in the proposed scheme, rather than 1-slot drifts in the existing works, we study the  $T$ -slot drifts for our scheme, which are then approximated by constructing a virtual sequence  $X'_{t+s}$  conditioned on the state of  $X_t$  in slot  $t$ .

*Step 2 - Property analysis of drifts:* Based on the approximation of  $T$ -slot drifts for FASA, we obtain the properties of the  $T$ -slot drifts required for guaranteeing the stability of the scheme.

*Step 3 - Stability analysis based on Lyapunov function:* The Lyapunov function defined in [13] is adopted for the proposed scheme. Then with the the properties obtained in Step 2, we show the geometrical ergodicity of  $X_t$  by analyzing the drifts of the Lyapunov function.

### Step 1 - Approximation of Drifts

In order to analyze the stability of FASA, we evaluate the change of  $X_t$  from slot  $t$  to slot  $t+T$ . Let  $\tilde{N}_t = \hat{N}_t - N_t$

denote the estimate error in slot  $t$ . Conditioned on the state of  $X_t$ , we define the  $T$ -slot drifts of  $N_t$ ,  $\hat{N}_t$ , and  $\tilde{N}_t$  as follows:

$$d_T(n, \hat{n}, k) = E[N_{t+T} - N_t | X_t = (n, \hat{n}, k)], \quad (21)$$

$$\hat{d}_T(n, \hat{n}, k) = E[\hat{N}_{t+T} - \hat{N}_t | X_t = (n, \hat{n}, k)], \quad (22)$$

$$\begin{aligned} \tilde{d}_T(n, \hat{n}, k) &= E[\tilde{N}_{t+T} - \tilde{N}_t | X_t = (n, \hat{n}, k)] \\ &= \hat{d}_T(n, \hat{n}, k) - d_T(n, \hat{n}, k). \end{aligned} \quad (23)$$

It is difficult to calculate the drifts defined above and we try to obtain the properties of them by introducing an approximate version of  $X_t$ . Let  $\{Z'_{t+s} : s \in \mathbb{Z}_+\}$  be a ternary independently and identically distributed (i.i.d.) random sequence, whose distribution is given by

$$\Pr(Z'_{t+s} = i) = q_i(\rho), \quad i = 0, 1, c,$$

where  $\rho = n/\hat{n}$ . Then we construct a virtual sequence  $X'_{t+s} = (N'_{t+s}, \hat{N}'_{t+s}, K'_{t+s})$  based on  $X_t$  and  $Z'_{t+s}$  as follows: when  $s = 0$ ,  $X'_t = X_t = (n, \hat{n}, k)$ ; when  $s > 0$ ,  $K'_{t+s}$ ,  $N'_{t+s}$ , and  $\hat{N}'_{t+s}$  are updated in a similar way as (18) - (20), respectively, with  $Z_t$  replaced with  $Z'_{t+s}$ . However, unlike the updates in  $X_t$ , we allow  $N'_{t+s}$  to be negative and  $\hat{N}'_{t+s}$  to be less than 1.

Obviously,  $X'_{t+s}$  is a Markov chain and its transition probabilities are fixed and determined by the state of  $X_t$ . We define its  $T$ -slot drifts  $d'_T(n, \hat{n}, k)$ ,  $\hat{d}'_T(n, \hat{n}, k)$ , and  $\tilde{d}'_T(n, \hat{n}, k)$  similarly to (21) - (23). When  $T$  is given, we show in Lemma 2 that the drifts of  $X'_{t+s}$  can be used to approximate the drifts of  $X_t$  from slot  $t$  to slot  $t+T$ , when either  $N_t$  or  $\hat{N}_t$  is sufficiently large.

**Lemma 2** *Given  $T > 0$  and  $\epsilon > 0$ , there exists some  $M > 0$ , such that if  $N_t = n \geq M$  or  $\hat{N}_t = \hat{n} \geq M$ , then*

$$|d_T(n, \hat{n}, k) - d'_T(n, \hat{n}, k)| \leq \epsilon, \quad (24)$$

$$|\hat{d}_T(n, \hat{n}, k) - \hat{d}'_T(n, \hat{n}, k)| \leq \epsilon. \quad (25)$$

*Proof:* See Appendix C.

According to Lemma 2, for given  $T$ , the differences between the  $T$ -slot drifts of  $X'_{t+s}$  and  $X_{t+s}$  can be made as close to zero as desired by letting  $n$  or  $\hat{n}$  be sufficiently large. Next, we evaluate the drifts of  $X'_{t+s}$  for obtaining the properties of drifts for FASA in Step 2.

1)  $d'_T(n, \hat{n}, k)$ : Since in the virtual sequence  $X'_{t+s}$ ,  $N'_{t+s}$  is allowed to be negative, we can easily have

$$\begin{aligned} d'_T(n, \hat{n}, k) &= E\left[\sum_{s=0}^{T-1} (A_{t+s} - Z'_{t+s}) | X'_t = (n, \hat{n}, k)\right] \\ &= T(\bar{\lambda} - \rho e^{-\rho}). \end{aligned}$$

2)  $\hat{d}'_T(n, \hat{n}, k)$ : In slot  $t+s$ , the update of  $\hat{N}'_{t+s}$  depends on both  $K'_{t+s}$  and  $Z'_{t+s}$ . Notice that the sequence  $K'_{t+s}$  is a Markov chain on a finite state space  $\mathbb{S}_K = \{-k_m, -k_m + 1, \dots, 0, \dots, k_m\}$ . Since the distribution of  $Z'_{t+s}$  is fixed, by showing the ergodicity of  $K'_{t+s}$ , we are able to approximate the  $T$ -slot drift by analyzing the stationary behavior of  $K'_{t+s}$ . Specifically, the transition of  $K'_{t+s}$  depends on the value of

$Z'_{t+s}$  and the 1-step transition probabilities is given by

$$p_{kj}(\rho) = \begin{cases} q_0(\rho), & \text{if } k \leq 0, j = \max\{k-1, -k_m\}, \\ q_0(\rho), & \text{if } k > 0, j = -1, \\ q_1(\rho), & \text{if } j = 0, \\ q_c(\rho), & \text{if } k < 0, j = 1, \\ q_c(\rho), & \text{if } k \geq 0, j = \min\{k+1, k_m\}, \\ 0, & \text{else.} \end{cases}$$

It is easy to verify that when  $\rho > 0$ ,  $K'_{t+s}$  is irreducible and aperiodic. Thus,  $K'_{t+s}$  is ergodic and there is a unique stationary distribution. Now we study the stationary distribution of  $K'_{t+s}$  and the drift of  $\hat{N}'_{t+s}$  in the steady state. Define a  $1 \times (2k_m + 1)$  vector as follows:

$$\boldsymbol{\pi}(\rho) = [\pi_{-k_m}(\rho), \pi_{-k_m+1}(\rho), \dots, \pi_0(\rho), \dots, \pi_{k_m}(\rho)],$$

where the elements are given by

$$\pi_k(\rho) = \begin{cases} q_0^{k_m}(\rho), & \text{if } k = -k_m, \\ q_0^{|k|}(\rho)[1 - q_0(\rho)], & \text{if } -k_m + 1 \leq k \leq -1, \\ q_1(\rho), & \text{if } k = 0, \\ q_c^k(\rho)[1 - q_c(\rho)], & \text{if } 1 \leq k \leq k_m - 1, \\ q_c^{k_m}(\rho), & \text{if } k = k_m. \end{cases}$$

It can be verified that  $\sum_{j=-k_m}^{k_m} \pi_j(\rho) = 1$  and  $\pi_j(\rho) = \sum_{k=-k_m}^{k_m} \pi_k(\rho) p_{kj}(\rho)$  for all  $k \in \mathbb{S}_K$ . Hence,  $\boldsymbol{\pi}(\rho)$  is the stationary distribution of  $K'_{t+s}$ . Using the expression of  $\boldsymbol{\pi}(\rho)$ , we can verify that  $\varphi(\rho)$  defined in (14) represents the stationary drift of  $\hat{N}'_{t+s}$ , which is the 1-slot drift of  $\hat{N}'_{t+s}$  when  $K'_{t+s}$  is in the steady state. Consequently, with the ergodicity of  $K'_{t+s}$ , we have

$$\lim_{T \rightarrow \infty} \frac{1}{T} \hat{d}'_T(n, \hat{n}, k) = \varphi(\rho).$$

Moreover, in order to use Lemma 2, we expect to find a common  $T$  such that (24) and (25) hold for some given  $\epsilon$  and for all  $(n, \hat{n}, k) \in \mathbb{S}_X$ , which requires the uniform convergence of  $\frac{1}{T} \hat{d}'_T(n, \hat{n}, k)$ . In fact, by analyzing the evolution of  $K'_{t+s}$ , we can show that  $\frac{1}{T} \hat{d}'_T(n, \hat{n}, k)$  converges uniformly in  $(n, \hat{n}, k)$ . First, by multiplying the transition probability matrix  $k_m$  times or analyzing the event that  $K'_{t+k_m} = j$ , we can see that for any  $k, j \in \mathbb{S}_K$ , the  $k_m$ -step transition probability  $p_{k,j}^{(k_m)}(\rho) = \pi_j(\rho)$ . For example, for any  $k \in \mathbb{S}_K$  and  $j \in (-k_m, 0)$ ,  $K'_{t+k_m} = j$  holds if and only if  $Z'_{t+s} = 0$  for all  $s = k_m - 1, k_m - 2, \dots, k_m - j + 1$ , while  $Z'_{t+k_m-j} \neq 0$ , so  $p_{k,j}^{(k_m)}(\rho) = q_0^{|j|}(\rho)[1 - q_0(\rho)] = \pi_j(\rho)$ . Consequently, for any state  $k \in \mathbb{S}_K$ , we have

$$\Pr(K'_{t+s} = j) = \pi_j(\rho), \quad s \geq k_m,$$

and thus when  $s \geq k_m$  the drift of  $\hat{N}'_{t+s}$  in each slot is exactly  $\varphi(\rho)$ . Then, with the fact that for any  $(n, \hat{n}, k) \in \mathbb{S}_X$ ,

$$\begin{aligned} & \left| \hat{d}'_{k_m-1}(n, \hat{n}, k) - (k_m - 1)\varphi(\rho) \right| \\ & \leq (k_m - 1) \left[ \frac{1}{e-2} + k_m^\nu \max\{h_0(\nu), h_c(\nu)\} \right], \end{aligned}$$

we know that as  $T$  tends to infinity,  $\frac{1}{T} \hat{d}'_T(n, \hat{n}, k)$  converges to  $\varphi(\rho)$  uniformly in  $(n, \hat{n}, k)$ . Thus, the difference between  $\frac{1}{T} \hat{d}'_T(n, \hat{n}, k)$  and  $\varphi(\rho)$  can be made as close to zero as desired by choosing a common  $T$  for all  $(n, \hat{n}, k) \in \mathbb{S}_X$ .

3)  $\tilde{d}'_T(n, \hat{n}, k)$ : Since  $\tilde{d}'_T(n, \hat{n}, k) = \hat{d}'_T(n, \hat{n}, k) - d'_T(n, \hat{n}, k)$ , we introduce the following function to approximate  $\frac{1}{T} \tilde{d}'_T(n, \hat{n}, k)$ :

$$\begin{aligned} \psi(\rho, \bar{\lambda}) &= \varphi(\rho) - (\bar{\lambda} - \rho e^{-\rho}) \\ &= \left[ \frac{1}{e-2} + h_c(\nu)\mu(\nu, q_c(\rho), k_m) \right] (1 - e^{-\rho} - \rho e^{-\rho}) \\ &\quad - [1 + h_0(\nu)\mu(\nu, q_0(\rho), k_m)] e^{-\rho} + \rho e^{-\rho} - \bar{\lambda}. \end{aligned}$$

With the uniform convergence of  $\frac{1}{T} \hat{d}'_T(n, \hat{n}, k)$ , we know that for any given  $\bar{\lambda} > 0$ , as  $T \rightarrow \infty$ ,

$$\frac{1}{T} \tilde{d}'_T(n, \hat{n}, k) \rightarrow \psi(\rho, \bar{\lambda})$$

uniformly in  $(n, \hat{n}, k)$ .

### Step 2 - Property Analysis of Drifts

The evolution of estimate error  $\tilde{N}_t$  is critical for showing the stability of the scheme. We first show in Lemma 3 that the approximate drift  $\psi(\rho, \bar{\lambda})$  has the same properties as those for the 1-slot drift of PB-ALOHA and then present the required properties of  $T$ -slot drifts of FASA in Lemma 4.

**Lemma 3** Given  $k_m$  as a positive integer,  $\psi(\rho, \bar{\lambda})$  has the following properties:

- a) For any  $\bar{\lambda}$ , the function  $\psi(\rho, \bar{\lambda})$  is strictly increasing in  $\rho$ .
- b) For any  $\bar{\lambda} \in (0, e^{-1}]$ , there exists a unique  $\rho = \omega(\bar{\lambda}) \in (0, 1]$ , such that  $\psi(\rho, \bar{\lambda}) = 0$ .
- c) If  $\bar{\lambda} \in (0, e^{-1})$ , then  $\omega(\bar{\lambda})e^{-\omega(\bar{\lambda})} > \bar{\lambda}$ .

*Proof:* See Appendix D.

Let  $\beta = \omega(\bar{\lambda})$  denote the root of  $\psi(\rho, \bar{\lambda}) = 0$  for given  $\bar{\lambda}$ . Similarly to the method in [13], we partition the state space into the following four parts:

$$S_{\gamma, M} = \{(n, \hat{n}, k) : n \geq M \text{ or } \hat{n} \geq M,$$

$$\beta - \gamma \leq \frac{n}{\hat{n}} \leq 1 + \gamma, k \in \mathbb{S}_K\},$$

$$R_{\gamma, M}^- = \{(n, \hat{n}, k) : \hat{n} \geq M, \frac{n}{\hat{n}} < \beta - \gamma, k \in \mathbb{S}_K\},$$

$$R_{\gamma, M}^+ = \{(n, \hat{n}, k) : n \geq M, \frac{n}{\hat{n}} > 1 + \gamma, k \in \mathbb{S}_K\},$$

$$Q_M = \{(n, \hat{n}, k) : n < M, \hat{n} < M, k \in \mathbb{S}_K\},$$

and let  $R_{\gamma, M} = R_{\gamma, M}^- \cup R_{\gamma, M}^+$ .

With Lemmas 2 and 3, we present the properties of the  $T$ -slot drifts in these regions in the following lemma.

**Lemma 4** There exist some  $\gamma > 0$ ,  $\delta > 0$ ,  $T > 0$ , and  $M > 0$ , such that  $5\gamma < \beta$  and

$$d_T(n, \hat{n}, k) \leq -T\delta, \quad \forall (n, \hat{n}, k) \in S_{5\gamma, M}, \quad (26)$$

$$\tilde{d}_T(n, \hat{n}, k) \leq -T\delta, \quad \forall (n, \hat{n}, k) \in R_{\gamma, M}^-, \quad (27)$$

$$\tilde{d}_T(n, \hat{n}, k) \geq T\delta, \quad \forall (n, \hat{n}, k) \in R_{\gamma, M}^+. \quad (28)$$

*Proof:* See Appendix E.

Intuitively, according to Lemma 4, when the estimate  $\hat{N}_t$  is close enough to  $N_t$ , positive number of devices will access successfully and leave the network in the following  $T$  slots. On the other hand, the deviation of the estimate  $\hat{N}_t$  from  $N_t$  is expected to decrease when it is larger than a certain threshold. These properties guarantee the stability of FASA, as presented in Step 3.



### Step 3 - Stability Analysis Based on Lyapunov Function

Lemma 4 shows that with a sufficiently large  $T$ , the  $T$ -slot drifts have similar properties to the drifts of PB-ALOHA. Hence, when observing the system every  $T$  slots, the Lyapunov function based method for PB-ALOHA can be used for analyzing the stability of FASA. Next, we provide an outline of using the Lyapunov function based method to prove the stability of FASA. For more details about this method, it is recommended to refer to [13].

Assume that  $T$ ,  $M$ ,  $\gamma$ , and  $\delta$  are fixed and that inequations (26) - (28) hold. We use the Lyapunov function defined in [13]:

$$V(n, \hat{n}, k) = \max \left\{ n, \frac{1+3\gamma}{3\gamma}(n - \hat{n}), \frac{\beta - 3\gamma}{1 - \beta + 3\gamma}(\hat{n} - n) \right\} \\ = \begin{cases} n, & \text{if } (n, \hat{n}, k) \in S_{3\gamma, M}, \\ \frac{1+3\gamma}{3\gamma}(n - \hat{n}), & \text{if } (n, \hat{n}, k) \in R_{3\gamma, M}^+, \\ \frac{\beta - 3\gamma}{1 - \beta + 3\gamma}(\hat{n} - n), & \text{if } (n, \hat{n}, k) \in R_{3\gamma, M}^-. \end{cases} \quad (29)$$

We will show that if  $J$  is sufficiently large, there exists some  $\Delta > 0$  such that

$$E[V(N_{t+JT}, \hat{N}_{t+JT}, K_{t+JT}) - V(N_t, \hat{N}_t, K_t) + \Delta; (N_t, \hat{N}_t, K_t) \notin Q_{M+(JT)^2} | \mathcal{F}_t] \leq 0, \quad (30)$$

where  $\mathcal{F}_t$  is the  $\sigma$ -field generated by  $\{A_{s-1}, N_s, \hat{N}_s, K_s : s \leq t\}$  and for random variable  $X$  and event  $A$ , the notation  $E[X; A | \mathcal{F}]$  stands for  $E[XI(A) | \mathcal{F}]$ .

For given  $t \geq 0$  and integer  $J$ , let

$$\tau_J = \min\{j \geq 0 : \sum_{s=0}^{jT} A_{t+s} \geq JT\}.$$

Similarly to [13], we then analyze the drift of the Lyapunov function by considering the unlikely event  $\{\tau_J \leq J\}$  and likely event  $\{\tau_J > J\}$  separately.

Using Chernoff bound [28], we can show that the following results also hold for interrupted Poisson process:

$$\lim_{J \rightarrow \infty} \Pr(\tau_J \leq J) = 0, \quad (31)$$

$$\lim_{J \rightarrow \infty} E \left[ l_1 JT + l_2 \sum_{s=0}^{JT} A_{t+s}; \tau_J \leq J \right] = 0, \quad (32)$$

where  $l_1, l_2$  are arbitrary given constants. Thus, for any  $(n, \hat{n}, k) \in \mathbb{S}_X$ , as  $J \rightarrow \infty$ , we have

$$E[|V(N_{t+JT}, \hat{N}_{t+JT}, K_{t+JT}) - V(N_t, \hat{N}_t, K_t)|; \tau_J \leq J | X_t = (n, \hat{n}, k)] \rightarrow 0, \quad (33)$$

implying that this expectation can be made as close to 0 as desired by choosing a sufficiently large  $J$ .

Now consider the event  $\tau_J > J$ . Based on the value of  $X_t = (n, \hat{n}, k)$ , we study the drift of the Lyapunov function in the following five cases:

- a)  $(n, \hat{n}, k) \in S_{2\gamma, M+(JT)^2}$ ;
- b)  $(n, \hat{n}, k) \in R_{4\gamma, M+(JT)^2}^+$ ;
- c)  $(n, \hat{n}, k) \in R_{4\gamma, M+(JT)^2}^-$ ;
- d)  $(n, \hat{n}, k) \in R_{4\gamma, M+(JT)^2} \cap R_{2\gamma, M+(JT)^2}^+$ ;
- f)  $(n, \hat{n}, k) \in R_{4\gamma, M+(JT)^2} \cap R_{2\gamma, M+(JT)^2}^-$ .

In any of these cases, following the approach in [13], we can show that when  $J$  is sufficiently large, there exists some  $\Delta > 0$ , such that inequation (30) holds.

Take case a) as an example. According to Lemma 3.4 in [13], if  $X_t = (n, \hat{n}, k) \in S_{2\gamma, M+(JT)^2}$ , then we choose a sufficiently large  $J$ , such that  $(N_{t+s}, \hat{N}_{t+s}, K_{t+s}) \in S_{3\gamma, M}$  for all  $s = 0, 1, \dots, JT$ , and

$$V(N_{t+jT}, \hat{N}_{t+jT}, K_{t+jT}) = N_{t+jT}, \quad \text{for all } j \in [0, J].$$

Thus, choosing a sufficiently large  $J$  such that  $\Pr(\tau_J > J) > 1/2$ , we have

$$E[V(N_{t+JT}, \hat{N}_{t+JT}, K_{t+JT}) - V(N_t, \hat{N}_t, K_t); \tau_J > J | X_t = (n, \hat{n}, k)] \\ = E[N_{t+JT} - N_t; \tau_J > J | X_t = (n, \hat{n}, k)] \\ = \sum_{j=0}^{J-1} E[d_T(N_{t+jT}, \hat{N}_{t+jT}, K_{t+jT}); \tau_J > J | X_t = (n, \hat{n}, k)] \\ \leq -\delta T J \Pr(\tau_J > J) \leq -\frac{\delta T J}{2}. \quad (34)$$

Combining (33) and (34), we know that there exists some  $J$  such that inequation (30) holds for some given  $\Delta > 0$ .

Now we are able to use the results about the hitting time bounds implied by drift analysis in [12]. Let

$$\Lambda = [M + (JT)^2] \max\{1, \frac{1+3\gamma}{3\gamma}, \frac{\beta - 3\gamma}{1 - \beta + 3\gamma}\}.$$

Note that for any  $K_t \in \mathbb{S}_K$ , whenever  $N_t \geq M + (JT)^2$  or  $\hat{N}_t \geq M + (JT)^2$ , (30) holds. According to Theorem 2.3 in [12], for any initiate state, the returning time  $\tau_\Lambda^* = \min\{t > 0 : V(N_t, \hat{N}_t, K_t) < \Lambda\}$  is exponential type, which implies that  $X_t$  is geometrically ergodic and concludes the proof of Theorem 2. ■

Similarly to the discussion in [13], from the proof of Theorem 2, we know that the proposed FASA scheme is stable under more general traffics, as long as the average arrival rate  $\bar{\lambda} < e^{-1}$  and the traffic model satisfies the conditions in (31) and (32).

## VI. SIMULATION RESULTS

In this section we evaluate the performance of the proposed scheme through simulation. We first examine the tracking performance and the effect of control parameters  $\nu$  and  $\eta$  on the access performance. We then study the access delay of the proposed scheme, including both the cases of single event and multiple events reporting.

We compare the performance of our FASA scheme, the ideal policy with perfect knowledge of backlog, PB-ALOHA [21], and  $Q^+$ -Algorithm [24]. With perfect knowledge of  $N_t$ , the ideal policy sets transmission probability at  $p_t = 1/N_t$  for  $N_t > 0$ . Thus, the ideal policy achieves the minimum access delay of S-ALOHA and serves as a benchmark in the comparison. For PB-ALOHA, we use the estimated arrival rate  $\hat{\lambda}_t = e^{-1}$ , as suggested in [13].  $Q^+$ -Algorithm belongs to the class of multiplicative schemes which is first proposed by Hajek and van Loon [22]. In  $Q^+$ -Algorithm,  $\hat{N}_t$  is updated as follows:

$$\hat{N}_{t+1} = \max\{1, [I(Z_t = 0)/\zeta_0 + I(Z_t = 1) + \zeta_c I(Z_t = c)]\hat{N}_t\},$$



where  $\zeta_0 = 2^{0.25} \approx 1.1892$  and  $\zeta_c = 2^{0.35} \approx 1.2746$  are suggested in [24] for optimal performance.

### A. Performance of tracking

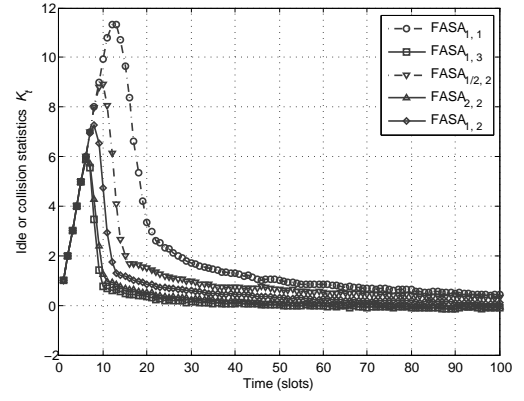
In order to gain more insights into the operation of the estimate schemes, we treat adaptive S-ALOHA schemes as dynamic systems and study their step responses, where the number of backlogged devices  $N_t = 0$  when  $t < 0$  and  $N_t = n$  for all  $t \geq 0$ . We examine the tracking performance of schemes for  $n = 500, 1000$ , and  $2000$ . For the sake of simplicity, we fix the value of  $k_m$  in FASA at 20, since the effect of  $k_m$  vanishes when it is large enough due to the exponential decay of distribution of  $K_{t,c}$ .

Before quantitative analysis, we first show the evolution of estimations under different conditions, which gives us some perceptual understanding about the behavior of these schemes.

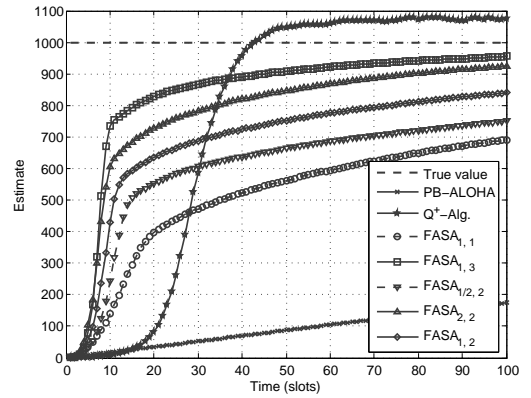
Fig. 3 shows the evolution of  $K_t$  and  $\hat{N}_t$  for different adaptive schemes, where the subscripts of FASA represents values of  $\eta$  and  $\nu$ . Average values of  $K_t$  and  $\hat{N}_t$  in each slot are obtained from 4000 independent experiments. It can be seen from this figure that, unlike the almost linearly increasing in PB-ALOHA, the estimate  $\hat{N}_t$  given by FASA increases slowly at the beginning, but speeds up due to the consecutive collisions, i.e., the increment of  $K_t$ . When the estimate gets close to the true value, success and idle slots occur and hence the increment of the estimate slows down. The estimate of Q<sup>+</sup>-Algorithm follows the same trend as FASA and speeds up even faster on average than FASA because of the exponentially increment. Though the drift of the estimate has the right direction, Q<sup>+</sup>-Algorithm turns out to be a bias estimate scheme, which will result in the suffering of the throughput at steady state. When comparing the curves of estimate for FASA with different parameters, we can see that with larger  $\eta$  or  $\nu$ , the estimate adjusts faster.

Fig. 4 shows the evolution of  $K_t$  and  $\hat{N}_t$  for different numbers of devices  $n$ . As shown in the figure, for larger  $n$ , there are more consecutive collisions at the beginning, which results in larger increment of estimate. After the peak point, the average value of  $K_t$  mainly depends on the offered load  $\rho_t = n/\hat{N}_t$  or the ratio of  $\hat{N}_t/n$ , so does the step size. For example, since  $\hat{N}_{10} \approx 300$  for  $n = 500$ , and  $\hat{N}_{30} \approx 1200$  for  $n = 2000$ , i.e., the ratios of  $\hat{N}_t/n$  in these slots for  $n = 500$  and  $1000$  are both about 0.6, they have almost the same average value of  $K_t$ , which is about 1. Though  $K_t$  decreases and FASA behaves like a fixed step size scheme as the estimate gets close to the true value, the value of  $K_t$  with larger  $n$  is larger than that with small  $n$ . Hence, we can expect that the time taken by FASA to catch up the true value or certain proportion of the true value will increase more slowly than linearly in  $n$ .

Because the access delay depends on the throughput, two throughput-oriented metrics are introduced to measure the response speed and stationary performance: 0%- $x$ % throughput rising time and stationary throughput. The 0%- $x$ % throughput rising time is defined as the time required for the expected throughput to rise from 0% to  $x$ % of the optimal value  $e^{-1}$ . For  $x = 10, 50$ , and  $90$ , they are equal to the time required for the estimated number of backlogged devices  $\hat{N}_t$  to rise from 0% to 20.45%, 37.34%, and 65.25% of the true value  $n$ , respectively. Stationary throughput is the average throughput



(a)  $K_t$



(b)  $\hat{N}_t$

Fig. 3. Evolution of estimation with different parameters ( $n = 1000$ ).

after the time that the expected throughput reaches 90% of the optimal value.

As shown in Table I, for the same  $x$ , the 0%- $x$ % rising time (unit: slot) of PB-ALOHA almost linearly increases in  $n$  and is much larger than that of Q<sup>+</sup>-Algorithm and FASA. For instance, when  $n = 1000$ , the 0%-50% rising time of FASA with  $\eta = 1$  and  $\nu = 2$  is about 1/20 of that of PB-ALOHA. Moreover, due to the aggressive update in FASA, the 0%- $x$ % rising time increases more slowly rather than linearly as the number of devices  $n$  increases, especially when  $x$  is small. Comparing the rising time of FASA with different  $\eta$  and  $\nu$ , we observe that the increment of  $\eta$  or  $\nu$  results in reduction of rising time. With multiplicative adjustment, it takes longer time than FASA for Q<sup>+</sup>-Algorithm to reach 10% of the optimal value but shorter time to increase the expected throughput from 10% to 90% of the optimal value. In addition, the increment of the rising time in Q<sup>+</sup>-Algorithm is tiny as  $n$  grows. However, aggressive adjustment of estimate usually results in large fluctuation at the steady state, and thus lower stationary throughput, which is shown in Table II. Thus, trade-off between the rising time and stationary throughput is necessary for choosing the values of parameters.

### B. Access delay

In this section, simulation results about the access delay of adaptive S-ALOHA schemes are presented. In some event-driven M2M applications, response can be taken with partial

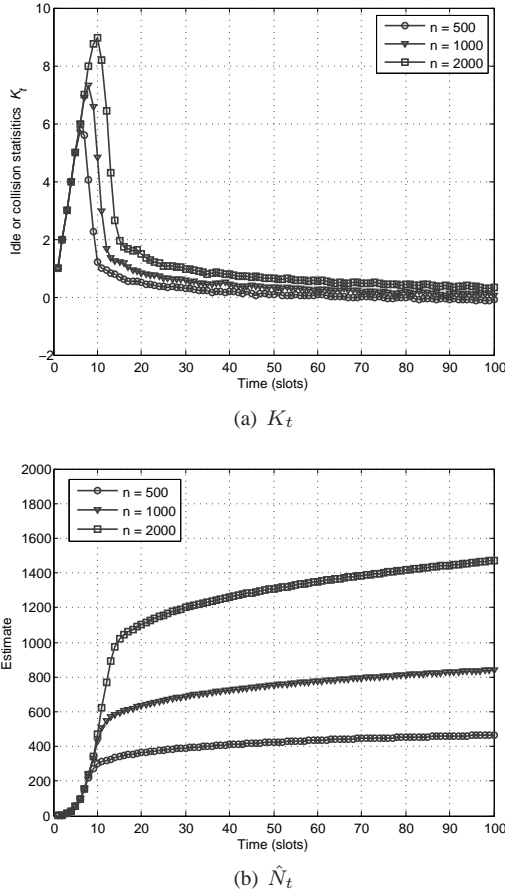


Fig. 4. Evolution of FASA for different  $n$  ( $\eta = 1, \nu = 2$ ).

TABLE I  
0% -  $x$ % THROUGHPUT RISING TIME

$n$	$x$	PB-ALOHA	$Q^+$ -Alg.	FASA <sub>1,1</sub>	FASA <sub>1,3</sub>	FASA <sub>1,2</sub>	FASA <sub>1/2,2</sub>	FASA <sub>2,2</sub>
500	10	59.8	21.0	9.1	6.0	7.1	8.1	5.0
	50	117.9	23.5	14.1	8.0	9.1	11.8	7.5
	90	313.1	27.5	43.9	14.5	21.3	31.9	15.6
1000	10	118.5	23.0	13.2	7.1	8.3	10.1	7.1
	50	237.0	26.6	21.6	9.4	12.0	15.0	9.0
	90	627.4	30.7	83.2	24.8	38.1	60.0	20.6
2000	10	236.0	26.0	18.3	8.1	10.1	12.1	8.1
	50	473.8	29.5	33.7	10.3	14.9	19.7	11.9
	90	1254.5	33.9	165.1	33.2	62.5	115.1	38.1

TABLE II  
STATIONARY THROUGHPUT

$n$	PB-ALOHA	$Q^+$ -Alg.	FASA <sub>1,1</sub>	FASA <sub>1,3</sub>	FASA <sub>1,2</sub>	FASA <sub>1/2,2</sub>	FASA <sub>2,2</sub>
500	0.3686	0.3529	0.3679	0.3656	0.3670	0.3676	0.3653
1000	0.3684	0.3521	0.3678	0.3663	0.3675	0.3681	0.3668
2000	0.3681	0.3523	0.3678	0.3674	0.3677	0.3678	0.3675

messages from the detecting devices and not all devices need to report an event. Thus, both the distribution of access delay for single event reporting and the long term average delay for repetitive event reporting are evaluated to study the performance of the proposed scheme.

1) *Single event reporting*: We focus on the scenarios where a large amount of devices are activated to report a single event and study the distribution of access delay of different adaptive schemes. Assume that  $N_0 = n$  devices are triggered at the

same time when an event is detected, and attempt to access the BS on the RACH. The scenarios with the number of active devices  $n = 500, 1000$ , and  $2000$  are studied.

Table III provides the  $y\%$  access delay (unit: slot), which is the access delay achieved by  $y\%$  of the active devices, and  $y$  is set to 10, 50, and 90. From Table III, we can see that the performance of the proposed FASA scheme is close to the benchmark with perfect information. For PB-ALOHA, it takes a long time to track the number of backlogged devices and few devices can access successfully during this period. For instance, the 10% delay is about two times of that for other schemes. For example, when  $n = 1000$ , the 10% delay of FASA<sub>1,2</sub> is 290.8 slots while it is 542.4 slots for PB-ALOHA. With multiplicative increment, the  $Q^+$ -Algorithm can track the number of backlogs in a short time because of the exponential increment due to the consecutive collision slots. However, it takes longer for all the devices to access the channel under  $Q^+$ -Algorithm than under FASA due to the large estimation fluctuations in  $Q^+$ -Algorithm. Comparing the access delay achieved by FASA with different  $\eta$  and  $\nu$ , we observe that the 10% access delay is slightly smaller for larger  $\eta$  or  $\nu$ , since they provide a quicker response ability. However, the larger fluctuation makes the 50% and 90% access delay for larger  $\eta$  and  $\nu$  close to, or even larger than that for smaller  $\eta$  and  $\nu$ .

TABLE III  
 $y\%$  ACCESS DELAY

$n$	$y$	Perf.Info.	PB-ALOHA	$Q^+$ -Alg.	FASA <sub>1,1</sub>	FASA <sub>1,3</sub>	FASA <sub>1,2</sub>	FASA <sub>1/2,2</sub>	FASA <sub>2,2</sub>
500	10	136.1	271.9	151.2	155.2	146.2	146.4	152.9	143.1
	50	680.6	822.3	712.5	702.7	694.4	691.3	700.6	691.0
	90	1223.1	1433.0	1282.9	1250.7	1252.1	1244.0	1254.2	1247.8
1000	10	267.3	542.4	298.6	306.4	285.9	290.8	303.3	282.9
	50	1351.9	1648.2	1426.0	1394.3	1384.7	1385.7	1385.8	1378.5
	90	2440.8	2871.4	2558.7	2496.0	2488.2	2484.1	2489.9	2483.5
2000	10	545.3	1083.8	585.1	614.5	563.7	568.6	592.4	564.3
	50	2712.1	3294.9	2855.8	2793.4	2749.2	2752.1	2779.1	2744.6
	90	4886.3	5745.8	5124.3	4983.1	4944.9	4944.0	4968.6	4934.5

2) *Repetitive events reporting with interrupted Poisson traffic*: The events happen sequentially in the real system and we now study the long term average delay of adaptive schemes under interrupted Poisson traffic with different arrival rates and bursty level. In addition to average delay, we also define the normalized divergence as follows to quantify the divergence from the theoretical optimum performance:

$$e(D) = \frac{D - D^*}{D^*},$$

where  $D$  is the average delay of a particular scheme and  $D^*$  is the theoretical optimal delay with perfect information. As pointed in the single event reporting case, the performance of FASA with different  $\eta$  and  $\nu$  are rather close. Thus, only the performance of FASA with  $\eta = 1$  and  $\nu = 2$  is presented here.

Fig. 5 compares the average delay and the normalized divergence of adaptive schemes under different arrival rates and fixed ON-probability  $\theta = 0.0001$ . From Fig. 5(a), we observe that both PB-ALOHA and FASA are stable when the average arrival rate  $\bar{\lambda} < e^{-1}$  and experience finite access delays. For  $Q^+$ -Algorithm, however, when the arrival rate is larger than about 0.352, the access delay grows unbounded, indicating that the algorithm with the given parameters is unstable for some  $\bar{\lambda} < e^{-1}$ . As pointed out in [12], the

parameters in  $Q^+$ -Algorithm should be carefully chosen to stabilize the scheme according to the value of  $\bar{\lambda}$ , which is not required in either PB-ALOHA or FASA. As shown in Fig. 5(b), when the arrival rate is close to zero, the divergence of  $Q^+$ -Algorithm and FASA are larger than that of PB-ALOHA due to the fluctuation of estimation, while all the delays are very small. As the average arrival rate increases, the divergence of FASA decreases and gets close to the optimal value, since the estimate error becomes relatively smaller compared to the increasing number of backlogged devices in the system. For  $\bar{\lambda}$  larger than 0.1, The divergence of FASA is about 2.5%, while it is about 22% for PB-ALOHA.

We point out that since the average access delay could be dominated by the time waiting in the system after the estimate catches up the true value, the improvement of performance by FASA does not seem to be significant from the average delay point of view. However, as discussed in the single event reporting cases, the 10% access delay can be improved significantly by FASA, which is very important to the event-driven M2M communications.

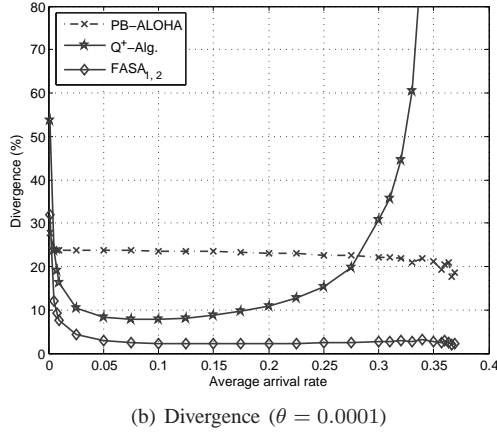
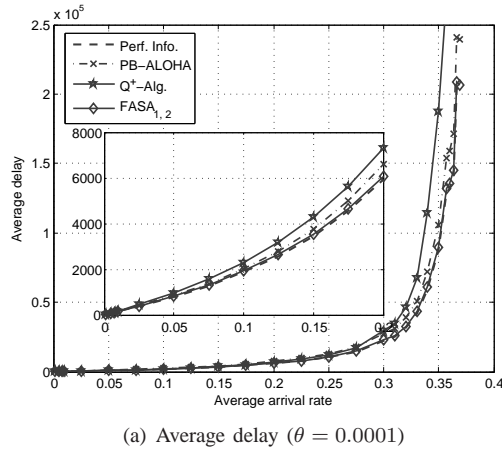


Fig. 5. Average access delay with different arrival rates.

In order to examine the impact of burstiness, we present in Fig. 6 the divergence of access delay versus variance of the arrival process  $\sigma_A^2$  for  $\bar{\lambda} = 0.05$  and  $0.35$ . For the light traffic scenarios with  $\bar{\lambda} = 0.05$ , when the bursty level is low, i.e., the variance is small, the access delay obtained from all these scheme are close to the optimal value. The reason is that there is usually only one device is triggered in one slot

when the estimate is usually set to 1 after several idle slots. As the traffic becomes more bursty, the divergence first increases owed to the rising time and fluctuation of estimate; and then the divergence of FASA and  $Q^+$ -Algorithm decreases for high bursty traffic because with aggressive update, they are able to track the status of the network quickly while the fluctuation become relatively smaller compared to the total number of backlogs. For the high traffic scenarios with  $\bar{\lambda} = 0.35$ , the divergences keep decreasing as the bursty level increase, while the proposed FASA scheme performs better than both PB-ALOHA and  $Q^+$ -Algorithm under high bursty traffic. Since the traffic in the event-driven M2M applications is bursty, we believe that our proposed scheme will perform well for these applications.

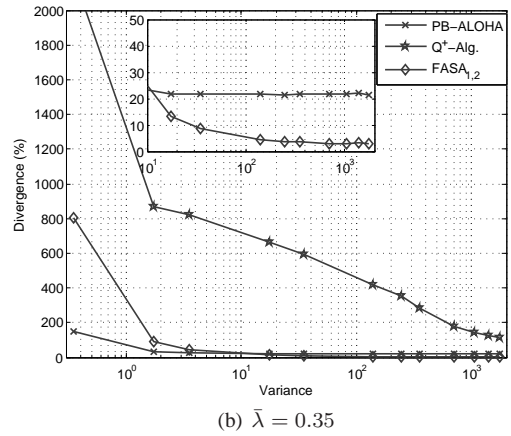
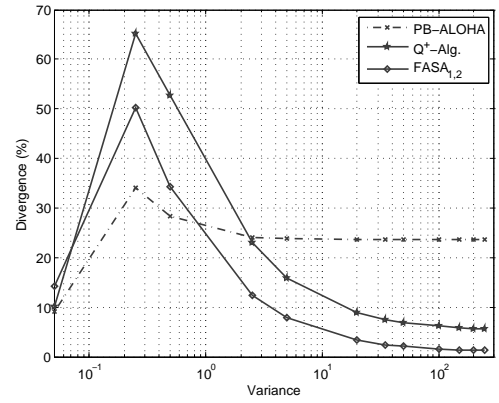


Fig. 6. Divergence of average access delay under different burstiness.

## VII. CONCLUSIONS

In this paper, we proposed a FASA scheme for event-driven M2M communications. By adjusting the estimate of the backlogs with statistics of consecutive idle and collision slots, a BS can track the number of backlogged devices more quickly. That is a main advantage compared to fixed step size additive schemes, e.g., PB-ALOHA. Moreover, we studied the stability of the proposed FASA under bursty traffic, which is modeled as an interrupted Poisson process. By analyzing the  $T$ -slot drifts of the FASA, we showed that without modifying the values of parameters, the proposed FASA scheme is stable for any average arrival rate less than  $e^{-1}$ , in the sense that



the system is geometrically ergodic. This property results in a much better long term average performance under heavy traffic loads, as compared with that of multiplicative schemes. In summary, the proposed scheme is an effective and stable S-ALOHA scheme and is suitable for the random access control of event-driven M2M communications as well as other systems characterized by bursty traffic.

#### APPENDIX A PROOF OF LEMMA 1

Recall that for  $\Delta n = \Delta \hat{n} = 0$ , it has been proved in [22] that when either the number of backlogged devices  $n$  or its estimate  $\hat{n}$  is sufficiently large, the idle and success probabilities (and hence as well as the collision probability) can be approximated as functions of the offered load  $\rho = n/\hat{n}$ . We generalize the results to any given  $\Delta n$  and  $\Delta \hat{n}$  by showing that as  $n$  or  $\hat{n}$  tends to infinity, the difference between the offered loads  $(n + \Delta n)/(\hat{n} + \Delta \hat{n})$  and  $\rho$  can be ignored and the distribution of access results can still be approximated by the same functions of  $\rho$ .

First, consider inequation (5), which is used for approximating the probability of  $Z_t = 0$ . Notice that

$$G^{(0)} = \left| e^{-\frac{n}{\hat{n}}} - \left(1 - \frac{1}{\hat{n} + \Delta \hat{n}}\right)^{n + \Delta n} \right| \\ \leq \underbrace{\left| e^{-\frac{n}{\hat{n}}} - e^{-\frac{n + \Delta n}{\hat{n} + \Delta \hat{n}}} \right|}_{G_1^{(0)}} + \underbrace{\left| e^{-\frac{n + \Delta n}{\hat{n} + \Delta \hat{n}}} - \left(1 - \frac{1}{\hat{n} + \Delta \hat{n}}\right)^{n + \Delta n} \right|}_{G_2^{(0)}}.$$

For  $G_2^{(0)}$ , according to Proposition 2.1 in [22], we know that there exists some  $M_2^{(0)} > 0$ , such that  $G_2^{(0)} \leq \epsilon/2$  for any  $(n, \hat{n}) \in H_{M_2^{(0)}}$ , where  $H_{M_2^{(0)}} = \{(n, \hat{n}) : n \geq M_2^{(0)} \text{ or } \hat{n} \geq M_2^{(0)}, n + \Delta n \geq 0, \hat{n} + \Delta \hat{n} \geq 1\}$ . Hence, we just need to show that  $G_1^{(0)} \leq \epsilon/2$  when either  $n$  or  $\hat{n}$  is sufficiently large.

Since  $\lim_{\rho \rightarrow \infty} e^{-\rho} = 0$ , there exists some  $\hat{\rho}^{(0)} > 1$  such that  $e^{-\rho} \leq \epsilon/4$  for any  $\rho > \hat{\rho}^{(0)}$ . Given a number  $\rho^{(0)} > \hat{\rho}^{(0)}$ , we analyze the value of  $G_1^{(0)}$  when  $n/\hat{n} > \rho^{(0)}$  and  $0 < n/\hat{n} \leq \rho^{(0)}$ , respectively.

0-a)  $n/\hat{n} > \rho^{(0)}$ : When  $n/\hat{n} > \rho^{(0)} > \hat{\rho}^{(0)}$ , it is easy to show that there exists some  $M_{1,1}^{(0)} > 0$ , such that if  $n \geq M_{1,1}^{(0)}$ , then  $(n + \Delta n)/(\hat{n} + \Delta \hat{n}) \geq (n - |\Delta n|)/(\hat{n} + |\Delta \hat{n}|) > \hat{\rho}^{(0)}$ . Hence, when  $n/\hat{n} > \rho^{(0)}$  and  $n \geq M_{1,1}^{(0)}$ , we have

$$G_1^{(0)} \leq e^{-\frac{n}{\hat{n}}} + e^{-\frac{n + \Delta n}{\hat{n} + \Delta \hat{n}}} \leq \epsilon/4 + \epsilon/4 = \epsilon/2.$$

0-b)  $0 < n/\hat{n} \leq \rho^{(0)}$ : Since  $n/\hat{n} \geq 0$ , we have

$$G_1^{(0)} = e^{-\frac{n}{\hat{n}}} \left| 1 - \exp\left(\frac{n}{\hat{n}} - \frac{n + \Delta n}{\hat{n} + \Delta \hat{n}}\right) \right| \\ \leq \left| 1 - \exp\left[\frac{n\Delta \hat{n} - \hat{n}\Delta n}{\hat{n}(\hat{n} + \Delta \hat{n})}\right] \right|.$$

When  $0 \leq n/\hat{n} \leq \rho^{(0)}$ , we have

$$\left| \frac{n\Delta \hat{n} - \hat{n}\Delta n}{\hat{n}(\hat{n} + \Delta \hat{n})} \right| \leq \frac{\rho^{(0)}|\Delta \hat{n}| + |\Delta n|}{(\hat{n} + \Delta \hat{n})} \rightarrow 0,$$

as  $\hat{n} \rightarrow \infty$ . Therefore,  $\lim_{\hat{n} \rightarrow \infty} \exp\left[\frac{n\Delta \hat{n} - \hat{n}\Delta n}{\hat{n}(\hat{n} + \Delta \hat{n})}\right] = 1$ , and hence, there exists some  $M_{1,2}^{(0)}$ , such that  $G_1^{(0)} \leq \epsilon/2$  for any  $(n, \hat{n})$  satisfying  $0 \leq n/\hat{n} \leq \rho^{(0)}$  and  $\hat{n} \geq M_{1,2}^{(0)}$ .

Consequently, combining all the cases analyzed above and letting  $M^{(0)} = \max\{M_2^{(0)}, M_{1,1}^{(0)}, \rho^{(0)}M_{1,2}^{(0)}\}$  follows that  $G^{(0)} \leq G_1^{(0)} + G_2^{(0)} \leq \epsilon$  for any  $(n, \hat{n}) \in H_{M^{(0)}}$ .

Next, we turn to the proof of inequation (6), which is used for approximating probability of  $Z_t = 1$ . Similarly to inequation (5),

$$G^{(1)} = \left| \frac{n}{\hat{n}} e^{-\frac{n}{\hat{n}}} - \frac{n + \Delta n}{\hat{n} + \Delta \hat{n}} \left(1 - \frac{1}{\hat{n} + \Delta \hat{n}}\right)^{n + \Delta n - 1} \right| \\ \leq \underbrace{\left| \frac{n}{\hat{n}} e^{-\frac{n}{\hat{n}}} - \frac{n + \Delta n}{\hat{n} + \Delta \hat{n}} e^{-\frac{n + \Delta n}{\hat{n} + \Delta \hat{n}}} \right|}_{G_1^{(1)}} \\ + \underbrace{\left| \frac{n + \Delta n}{\hat{n} + \Delta \hat{n}} e^{-\frac{n + \Delta n}{\hat{n} + \Delta \hat{n}}} - \frac{n + \Delta n}{\hat{n} + \Delta \hat{n}} \left(1 - \frac{1}{\hat{n} + \Delta \hat{n}}\right)^{n + \Delta n - 1} \right|}_{G_2^{(1)}}.$$

Using the result in [22], we know that there exists some  $M_2^{(1)} > 0$ , such that  $G_2^{(1)} \leq \epsilon/2$  for any  $(n, \hat{n}) \in H_{M_2^{(1)}}$  and we only need to show that  $G_1^{(1)} \leq \epsilon/2$  under certain conditions.

Since  $\lim_{\rho \rightarrow 0} \rho e^{-\rho} = \lim_{\rho \rightarrow \infty} \rho e^{-\rho} = 0$ , there exist some  $\hat{\rho}_1^{(1)}$  and  $\hat{\rho}_2^{(1)}$ , such that  $0 < \hat{\rho}_1^{(1)} < 1 < \hat{\rho}_2^{(1)}$ , and  $\rho e^{-\rho} \leq \epsilon/4$  for any  $\rho$  satisfying  $0 < \rho < \hat{\rho}_1^{(1)}$  or  $\rho > \hat{\rho}_2^{(1)}$ . Given  $\rho_1^{(1)} \in (0, \hat{\rho}_1^{(1)})$  and  $\rho_2^{(1)} \in (\hat{\rho}_2^{(1)}, \infty)$ , we study the following three cases based on the range of  $n/\hat{n}$ :

1-a)  $0 < n/\hat{n} < \rho_1^{(1)}$ : Similarly to the analysis of 0-a), since  $\rho_1^{(1)} < \hat{\rho}_1^{(1)}$ , there exists some  $M_{1,1}^{(1)} > 0$ , such that if  $\hat{n} > M_{1,1}^{(1)}$ , then  $(n + \Delta n)/(\hat{n} + \Delta \hat{n}) < \hat{\rho}_1^{(1)}$  and hence

$$G_1^{(1)} \leq \frac{n}{\hat{n}} e^{-\frac{n}{\hat{n}}} + \frac{n + \Delta n}{\hat{n} + \Delta \hat{n}} e^{-\frac{n + \Delta n}{\hat{n} + \Delta \hat{n}}} \leq \epsilon/2.$$

1-b)  $n/\hat{n} > \rho_2^{(1)}$ : Similarly to cases 0-a and 1-a), we can show that there exists some  $M_{1,2}^{(1)} > 0$ , such that if  $n > M_{1,2}^{(1)}$ , then  $G_1^{(1)} \leq \epsilon/2$ .

1-c)  $\rho_1^{(1)} \leq n/\hat{n} \leq \rho_2^{(1)}$ : When  $\rho_1^{(1)} \leq n/\hat{n} \leq \rho_2^{(1)}$ , we have

$$G_1^{(1)} = \frac{n}{\hat{n}} e^{-n/\hat{n}} \left| 1 - \frac{\hat{n}}{n} \cdot \frac{n + \Delta n}{\hat{n} + \Delta \hat{n}} \exp\left[\frac{n\Delta \hat{n} - \hat{n}\Delta n}{\hat{n}(\hat{n} + \Delta \hat{n})}\right] \right| \\ \leq \rho_2^{(1)} \left| 1 - \frac{1 + \Delta n/n}{1 + \Delta \hat{n}/\hat{n}} \exp\left[\frac{n\Delta \hat{n} - \hat{n}\Delta n}{\hat{n}(\hat{n} + \Delta \hat{n})}\right] \right|.$$

Since

$$\left| \frac{1 + \Delta n/n}{1 + \Delta \hat{n}/\hat{n}} - 1 \right| \leq \frac{|\Delta n|/\rho_1^{(1)} + |\Delta \hat{n}|}{|\hat{n} + \Delta \hat{n}|} \rightarrow 0,$$

as  $\hat{n} \rightarrow \infty$ , we have  $\lim_{\hat{n} \rightarrow \infty} \frac{1 + \Delta n/n}{1 + \Delta \hat{n}/\hat{n}} = 1$  and  $\lim_{\hat{n} \rightarrow \infty} \frac{1 + \Delta n/n}{1 + \Delta \hat{n}/\hat{n}} \exp\left[\frac{n\Delta \hat{n} - \hat{n}\Delta n}{\hat{n}(\hat{n} + \Delta \hat{n})}\right] = 1$ . Hence, there exists a  $M_{1,3}^{(1)}$  such that  $G_1^{(1)} \leq \epsilon/2$  for any  $(n, \hat{n})$  satisfying  $\rho_1^{(1)} \leq n/\hat{n} \leq \rho_2^{(1)}$  and  $\hat{n} \geq M_{1,3}^{(1)}$ .

Therefore,  $G^{(1)} \leq G_1^{(1)} + G_2^{(1)} \leq \epsilon$  for any  $(n, \hat{n}) \in H_{M^{(1)}}$ , where

$$M^{(1)} = \max\{M_2^{(1)}, M_{1,1}^{(1)}, M_{1,2}^{(1)}, \rho_2^{(1)}M_{1,3}^{(1)}\}.$$

Finally, from the above analysis, choosing  $M = \max\{M^{(0)}, M^{(1)}\}$ , we know that (5) and (6) hold for any  $(n, \hat{n}) \in H_M$  and this concludes the proof of Lemma 1.

APPENDIX B  
PROOF OF THEOREM 1

The proposition can be proved by calculating the derivative of  $\varphi(\rho)$ .

For given values of  $\nu$  and  $k_m$ , let

$$\begin{aligned}\varphi^{(0)}(\rho) &= |q_0(\rho)E[\Delta N_t|(0, n, \hat{n})]| \\ &= q_0(\rho)[1 + h_0(\nu)\mu(\nu, q_0(\rho), k_m)],\end{aligned}$$

and

$$\begin{aligned}\varphi^{(c)}(\rho) &= |q_c(\rho)E[\Delta N_t|(c, n, \hat{n})]| \\ &= q_c(\rho)[(e-2)^{-1} + h_c(\nu)\mu(\nu, q_c(\rho), k_m)].\end{aligned}$$

Then  $\varphi^{(0)}(1) = \varphi^{(c)}(1) = e^{-1} + \eta$  and  $\varphi(1) = -\varphi^{(0)}(1) + \varphi^{(c)}(1) = 0$ . Next, we claim that, for given  $\nu > 0$ ,  $\mu(\nu, q, k_m)$  defined in (12) is an increasing function of  $q$  ( $0 < q < 1$ ). This is because

$$\begin{aligned}\frac{\partial \mu}{\partial q} &= \sum_{k=1}^{k_m-1} k^\nu q^{k-2} [k(1-q) - 1] + (k_m - 1)k_m^\nu q^{k_m-2} \\ &= \sum_{k=1}^{k^*} k^\nu q^{k-2} [k(1-q) - 1] \\ &\quad + \sum_{k=k^*+1}^{k_m-1} k^\nu q^{k-2} [k(1-q) - 1] + (k_m - 1)k_m^\nu q^{k_m-2} \\ &> (k^*)^\nu \left\{ \sum_{k=1}^{k_m-1} q^{k-2} [k(1-q) - 1] + (k_m - 1)q^{k_m-2} \right\} \\ &= (k^*)^\nu \left[ \sum_{k=1}^{k_m} (k-1)q^{k-2} - \sum_{k=1}^{k_m-1} kq^{k-1} \right] = 0,\end{aligned}$$

where  $k^* = \lfloor (1-q)^{-1} \rfloor$  is the largest integer not greater than  $(1-q)^{-1}$ , and thus  $k^\nu \leq (k^*)^\nu$  if  $1 \leq k \leq k^*$  and  $k^\nu > (k^*)^\nu$  if  $k > k^*$ . In addition, the idle probability  $q_0(\rho) = e^{-\rho}$  is nonnegative and strictly decreasing in  $\rho$ . Hence,  $\mu(\nu, q_0(\rho), k_m)$  is strictly decreasing in  $\rho$  and  $\varphi^{(0)}(\rho)$  is a strictly decreasing function of  $\rho$ . On the other hand, since  $q_c(\rho) = 1 - e^{-\rho} - \rho e^{-\rho}$  is nonnegative and strictly increasing in  $\rho$ , we can similarly show that  $\varphi^{(c)}(\rho)$  is a strictly increasing function of  $\rho$ . Thus,  $\varphi(\rho) = -\varphi^{(0)}(\rho) + \varphi^{(c)}(\rho)$  is a strictly increasing function of  $\rho$ . Consequently,  $\varphi(\rho) < \varphi(1) = 0$  when  $0 < \rho < 1$  and  $\varphi(\rho) > \varphi(1) = 0$  when  $\rho > 1$ .

APPENDIX C  
PROOF OF LEMMA 2

Because of the similarity, we present a complete analysis of inequation (24), while discuss briefly about inequation (25) at the end. Recall that we assume the same realization of arrival process  $A_{t+s}$  for  $X_{t+s}$  and  $X'_{t+s}$ . In addition, given  $T$  and  $(n, \hat{n}, k)$ , the number of departures between slot  $t$  and  $t+T$  is bounded and the number of new arrivals can also be bounded with high probability. In order to use Lemma 1, we consider separately the events  $\sum_{s=0}^{T-1} A_{t+s} > B_A$  and  $\sum_{s=0}^{T-1} A_{t+s} \leq B_A$ .

$$1) \sum_{s=0}^{T-1} A_{t+s} > B_A$$

Using Chernoff bound, we can show that the probability  $\Pr(\sum_{s=0}^{T-1} A_{t+s} > B_A)$  decays exponentially as  $B_A$  grows to infinity. With the assumption of same realization of  $A_{t+s}$ ,

the differences between the  $T$ -slot drift of  $N_{t+s}$  and  $N'_{t+s}$  is bounded by  $T$ , i.e.,  $|d_T(n, \hat{n}, k) - d'_T(n, \hat{n}, k)| \leq T$ . Therefore, for any given  $\epsilon > 0$ , there exists some  $B_A > 0$  such that

$$|d_T(n, \hat{n}, k) - d'_T(n, \hat{n}, k)| \Pr\left(\sum_{s=0}^{T-1} A_{t+s} > B_A\right) \leq \epsilon/2. \quad (36)$$

2)  $\sum_{s=0}^{T-1} A_{t+s} \leq B_A$   
Let  $\mathbf{A}_{t,T} = (A_t, A_{t+1}, \dots, A_{t+T-1})$ ,  $\mathbf{Z}_{t,T} = (Z_t, Z_{t+1}, \dots, Z_{t+T-1})$ , and  $\mathbf{Z}'_{t,T} = (Z'_t, Z'_{t+1}, \dots, Z'_{t+T-1})$ . The set of possible values of  $(\mathbf{A}_{t,T}, \mathbf{Z}_{t,T})$  and  $(\mathbf{A}_{t,T}, \mathbf{Z}'_{t,T})$  are the same, denoted by  $\Omega$ . Since  $Z_{t+s}$  or  $Z'_{t+s}$  has three possible values and  $\sum_{s=0}^{T-1} A_{t+s} \leq B_A$ ,  $\Omega$  is a finite set. Each pair  $(\mathbf{a}, \mathbf{z}) = (a_0 \dots a_{T-1}, z_0 \dots z_{T-1}) \in \Omega$  results in corresponding drifts in both  $N_{t+s}$  and  $N'_{t+s}$ , denoted by  $\Delta N_{(\mathbf{a}, \mathbf{z}, n, \hat{n}, k)}$  and  $\Delta N'_{(\mathbf{a}, \mathbf{z}, n, \hat{n}, k)}$ , respectively.

One of the differences between  $X_{t+s}$  and  $X'_{t+s}$  is that we do not limit the values of  $N'_{t+s}$  and  $\hat{N}'_{t+s}$ , i.e.,  $N'_{t+s} < 0$  and  $\hat{N}'_{t+s} < 1$  are allowed in the virtual sequence, which is not the case in  $X_{t+s}$ . However, noticing the fact that when  $n \geq T$ , the drifts of  $N_{t+s}$  and  $N'_{t+s}$  are only decided by  $(\mathbf{a}, \mathbf{z})$ , we have  $\Delta N_{(\mathbf{a}, \mathbf{z}, n, \hat{n}, k)} = \Delta N'_{(\mathbf{a}, \mathbf{z}, n, \hat{n}, k)}$  for any  $(\mathbf{a}, \mathbf{z}) \in \Omega$  in this case. We first study the case where  $n \geq T$ , and analyze later the other cases where  $n$  is not large enough.

Conditioned on  $\sum_{s=0}^{T-1} A_{t+s} \leq B_A$ , we define the following probabilities:

$$\begin{aligned}f_{\mathbf{A}}(\mathbf{a}) &= \Pr(\mathbf{A}_{t,T} = \mathbf{a}), \\ f_{(\mathbf{A}, \mathbf{Z})|X}(\mathbf{a}, \mathbf{z}|n, \hat{n}, k) &= \Pr[(\mathbf{A}_{t,T}, \mathbf{Z}_{t,T}) = (\mathbf{a}, \mathbf{z}) | X_t = (n, \hat{n}, k)], \\ f_{(\mathbf{A}, \mathbf{Z}')|X'}(\mathbf{a}, \mathbf{z}|n, \hat{n}, k) &= \Pr[(\mathbf{A}_{t,T}, \mathbf{Z}'_{t,T}) = (\mathbf{a}, \mathbf{z}) | X'_t = (n, \hat{n}, k)], \\ f_{\mathbf{Z}|\mathbf{A}, X}(\mathbf{z}|\mathbf{a}, n, \hat{n}, k) &= \Pr[\mathbf{Z}_{t,T} = \mathbf{z} | \mathbf{A}_{t,T} = \mathbf{a}, X_t = (n, \hat{n}, k)], \\ f_{\mathbf{Z}'|\mathbf{A}, X'}(\mathbf{z}'|\mathbf{a}, n, \hat{n}, k) &= \Pr[\mathbf{Z}'_{t,T} = \mathbf{z}' | \mathbf{A}_{t,T} = \mathbf{a}, X'_t = (n, \hat{n}, k)].\end{aligned}$$

Thus, when  $n \geq T$ , we have

$$\begin{aligned}&|d_T(n, \hat{n}, k) - d'_T(n, \hat{n}, k)| \Pr\left(\sum_{s=0}^{T-1} A_{t+s} \leq B_A\right) \\ &\leq \left| \sum_{(\mathbf{a}, \mathbf{z}) \in \Omega} [f_{(\mathbf{A}, \mathbf{Z})|X}(\mathbf{a}, \mathbf{z}|n, \hat{n}, k) \right. \\ &\quad \left. - f_{(\mathbf{A}, \mathbf{Z}')|X'}(\mathbf{a}, \mathbf{z}'|n, \hat{n}, k)] \Delta N_{(\mathbf{a}, \mathbf{z}, n, \hat{n}, k)} \right| \\ &= \left| \sum_{(\mathbf{a}, \mathbf{z}) \in \Omega} [f_{\mathbf{Z}|\mathbf{A}, X}(\mathbf{z}|\mathbf{a}, n, \hat{n}, k) \right. \\ &\quad \left. - f_{\mathbf{Z}'|\mathbf{A}, X'}(\mathbf{z}'|\mathbf{a}, n, \hat{n}, k)] f_{\mathbf{A}}(\mathbf{a}) \Delta N_{(\mathbf{a}, \mathbf{z}, n, \hat{n}, k)} \right| \\ &\leq C \sum_{(\mathbf{a}, \mathbf{z}) \in \Omega} |f_{\mathbf{Z}|\mathbf{A}, X}(\mathbf{z}|\mathbf{a}, n, \hat{n}, k) - f_{\mathbf{Z}'|\mathbf{A}, X'}(\mathbf{z}'|\mathbf{a}, n, \hat{n}, k)|,\end{aligned}$$

where  $C$  is the maximum value of  $|f_{\mathbf{A}}(\mathbf{a}) \Delta N_{(\mathbf{a}, \mathbf{z}, n, \hat{n}, k)}|$  for all  $(\mathbf{a}, \mathbf{z}) \in \Omega$  and  $k \in \{-k_m, -k_m + 1, \dots, 0, \dots, k_m - 1, k_m\}$ .

Letting

$$\begin{aligned}g_{Z_s}(z_s|\mathbf{a}, z_0 \dots z_{s-1}) &= \Pr[Z_{t+s} = z_s | X_t = (n, \hat{n}, k), \\ &\quad \mathbf{A}_{t,T} = \mathbf{a}, Z_t \dots Z_{t+s-1} = z_0 \dots z_{s-1}], \\ g_{Z'_s}(z'_s|\mathbf{a}, z_t \dots z_{t+s-1}) &= \Pr[Z'_{t+s} = z'_s | X'_t = (n, \hat{n}, k), \\ &\quad \mathbf{A}_{t,T} = \mathbf{a}, Z'_t \dots Z'_{t+s-1} = z_0 \dots z_{s-1}],\end{aligned}$$

we have

$$\begin{aligned}
& f_{\mathbf{Z}|\mathbf{A},X}(\mathbf{z}|\mathbf{a},n,\hat{n},k) \\
&= \Pr(Z_t = z_0) \prod_{s=1}^{T-1} g_{Z_s}(z_s|\mathbf{a},z_0 \dots z_{s-1}) \\
&= \prod_{s=0}^{T-1} \Pr[Z_{t+s} = z_s | (N_{t+s}, \hat{N}_{t+s}) = (n_s, \hat{n}_s)_{\mathbf{a},\mathbf{z},n,\hat{n},k}], \\
& f_{\mathbf{Z}'|\mathbf{A},X}(\mathbf{z}|\mathbf{a},n,\hat{n},k) \\
&= \Pr(Z'_t = z_0) \prod_{s=1}^{T-1} g_{Z'_s}(z_s|\mathbf{a},z_0 \dots z_{s-1}) \\
&= \prod_{s=0}^{T-1} \Pr[Z'_{t+s} = z_s | (N_{t+s}, \hat{N}'_{t+s}) = (n_s, \hat{n}_s)_{\mathbf{a},\mathbf{z},n,\hat{n},k}],
\end{aligned}$$

where  $(n_s, \hat{n}_s)_{\mathbf{a},\mathbf{z},n,\hat{n},k}$  is the number of backlogged devices and its estimate in slot  $t+s$ , given the initial state  $X_t = X'_t = (n, \hat{n}, k)$ ,  $\mathbf{A}_{t,T} = \mathbf{a}$ , and  $\mathbf{Z}_{t,T} = \mathbf{Z}'_{t,T} = \mathbf{z}$ .

According to the construction of  $X'_{t+s}$ , the distribution of  $Z'_{t+s}$  is fixed, i.e.,  $\Pr(Z'_{t+s} = i) = q_i(\rho)$ ; while for  $X_{t+s}$ , there are finite possible values of  $N_{t+s} - n$  and  $\hat{N}_{t+s} - \hat{n}$ , for all  $s = 1, 2, \dots, T-1$ , since the initial value of  $K_t$  is from a finite set. Therefore, using Lemma 1, we know that there exists some sufficiently large  $M'_1$ , such that if  $n \geq M'_1$  or  $\hat{n} \geq M'_1$ , then for all  $s = 1, 2, \dots, T-1$ , and  $(\mathbf{a}, \mathbf{z}) \in \Omega$ ,

$$\begin{aligned}
& \left| \Pr[Z_{t+s} = z_s | (N_{t+s}, \hat{N}_{t+s}) = (n_s, \hat{n}_s)_{\mathbf{a},\mathbf{z},n,\hat{n},k}] \right. \\
& \left. - \Pr[Z'_{t+s} = z_s | (N_{t+s}, \hat{N}'_{t+s}) = (n_s, \hat{n}_s)_{\mathbf{a},\mathbf{z},n,\hat{n},k}] \right| \leq \frac{\epsilon}{2TC|\Omega|},
\end{aligned}$$

where  $|\Omega|$  is the number of elements in  $\Omega$ . Hence

$$|f_{\mathbf{Z}|\mathbf{A},X}(\mathbf{z}|\mathbf{a},n,\hat{n},k) - f_{\mathbf{Z}'|\mathbf{A},X}(\mathbf{z}|\mathbf{a},n,\hat{n},k)| \leq \frac{\epsilon}{2C|\Omega|},$$

and thus,

$$|d_T(n, \hat{n}, k) - d'_T(n, \hat{n}, k)| \Pr\left(\sum_{s=0}^{T-1} A_{t+s} \leq B_A\right) \leq \epsilon/2. \quad (37)$$

Therefore, combining (37) with (36) implies that (24) holds when  $n \geq T$ , and either  $n \geq M'_1$  or  $\hat{n} \geq M'_1$ .

Now consider the cases where  $n < T$ . In these cases,  $(\mathbf{A}_{t,T}, \mathbf{Z}_{t,T}) = (\mathbf{a}, \mathbf{z})$  is an impossible event for some  $(\mathbf{a}, \mathbf{z}) \in \Omega$ , if it results in some  $N_{t+s} < 0$ . For these  $(\mathbf{a}, \mathbf{z})$ , we set  $\Delta N_{(\mathbf{a},\mathbf{z},n,\hat{n},k)} = \Delta N'_{(\mathbf{a},\mathbf{z},n,\hat{n},k)}$ , which does not affect the calculation of drifts since the probabilities of these events in  $X_{t+s}$  are zero. Notice that for these  $(\mathbf{a}, \mathbf{z})$ , there is at least one component of  $\mathbf{z}$  equal to 1. Because  $\lim_{\rho \rightarrow 0} \rho e^{-\rho} = 0$ , we can find a number  $\rho_1 \in (0, 1)$ , such that  $\rho e^{-\rho} \leq \epsilon/(2TC|\Omega|)$  for all  $\rho = n/\hat{n} \leq \rho_1$ . Then the difference between  $f_{\mathbf{Z}|\mathbf{A},X}(\mathbf{z}|\mathbf{a},n,\hat{n},k)$  and  $f_{\mathbf{Z}'|\mathbf{A},X}(\mathbf{z}|\mathbf{a},n,\hat{n},k)$  is bounded by  $\epsilon/(2TC|\Omega|)$  and the same conclusion holds when  $n < T$  and  $n/\hat{n} \leq \rho_1$ .

Consequently, from the analysis above, we know that inequation (24) holds when  $n \geq M_1$  or  $\hat{n} \geq M_1$ , where  $M_1 = \max\{T/\rho_1, M'_1/\rho_1\}$ .

Finally, we discuss about inequation (25). For given  $(\mathbf{a}, \mathbf{z}) \in \Omega$ , denote the corresponding  $T$ -slot drifts of  $\hat{N}_{t+s}$  and  $\hat{N}'_{t+s}$  by  $\Delta \hat{N}_{(\mathbf{a},\mathbf{z},n,\hat{n},k)}$  and  $\Delta \hat{N}'_{(\mathbf{a},\mathbf{z},n,\hat{n},k)}$ , respectively. Note that for inequation (24), we do not specially treat the

case where  $\hat{n}$  is not large enough (but  $n$  is large enough), since we still have  $\Delta \hat{N}_{(\mathbf{a},\mathbf{z},n,\hat{n},k)} = \Delta \hat{N}'_{(\mathbf{a},\mathbf{z},n,\hat{n},k)}$  and the update of FASA guarantees that  $\hat{N}_{t+s} \geq 1$ . However,  $\Delta \hat{N}_{(\mathbf{a},\mathbf{z},n,\hat{n},k)} \neq \Delta \hat{N}'_{(\mathbf{a},\mathbf{z},n,\hat{n},k)}$  may occur for some  $(\mathbf{a}, \mathbf{z})$  when  $\hat{n}$  is small. Since for all  $(\mathbf{a}, \mathbf{z}) \in \Omega$ , both  $\Delta \hat{N}_{(\mathbf{a},\mathbf{z},n,\hat{n},k)}$  and  $\Delta \hat{N}'_{(\mathbf{a},\mathbf{z},n,\hat{n},k)}$  are bounded uniformly in  $(n, \hat{n}, k)$ , the impact of  $\Delta \hat{N}_{(\mathbf{a},\mathbf{z},n,\hat{n},k)} \neq \Delta \hat{N}'_{(\mathbf{a},\mathbf{z},n,\hat{n},k)}$  can be made ignorable by making the probability of this event as close to zero as possible with the fact that  $\lim_{\rho \rightarrow \infty} e^{-\rho} = 0$ . Hence, similarly to inequation (24), we can analyze the following three cases to obtain the threshold for inequation (25):

- a)  $n \geq T, \hat{n} \geq T[1 + k_m^\nu h_0(\nu)] + 1$ ;
- b)  $n < T, \hat{n} \geq T[1 + k_m^\nu h_0(\nu)] + 1$ ;
- c)  $n \geq T, \hat{n} < T[1 + k_m^\nu h_0(\nu)] + 1$ .

Therefore, the proof of this lemma can be concluded by choosing  $M$  as the larger threshold for inequation (24) and inequation (25).

#### APPENDIX D PROOF OF LEMMA 3

a) Let

$$\begin{aligned}
\psi^{(0)}(\rho, \bar{\lambda}) &= -e^{-\rho} + \frac{1}{e-2}(1 - e^{-\rho} - \rho e^{-\rho}), \\
\psi^{(1)}(\rho, \bar{\lambda}) &= \rho e^{-\rho} - \bar{\lambda}, \\
\psi^{(2)}(\rho, \bar{\lambda}) &= -e^{-\rho} h_0(\nu) \mu(\nu, q_0(\rho), k_m) \\
&\quad + (1 - e^{-\rho} - \rho e^{-\rho}) h_c(\nu) \mu(\nu, q_c(\rho), k_m).
\end{aligned}$$

Then,  $\psi = \psi^{(0)} + \psi^{(1)} + \psi^{(2)}$ .

First, because

$$\frac{\partial(\psi^{(0)} + \psi^{(1)})}{\partial \rho} = e^{-\rho} + \frac{\rho e^{-\rho}}{e-2} + e^{-\rho} - \rho e^{-\rho} > 0,$$

$\psi^{(0)} + \psi^{(1)}$  is strictly increasing in  $\rho$ .

In addition, we have shown in Appendix B that  $\mu(\nu, q_0(\rho), k_m)$  is strictly decreasing in  $\rho$  while  $\mu(\nu, q_c(\rho), k_m)$  is strictly increasing in  $\rho$ . Thus it is easy to verify that  $\psi^{(2)}(\rho, \bar{\lambda})$  is strictly increasing in  $\rho$ . Consequently,  $\psi = \psi^{(0)} + \psi^{(1)} + \psi^{(2)}$  is strictly increasing in  $\rho$ .

b) For any  $\bar{\lambda} \in (0, e^{-1}]$ , we have  $\psi(1, \bar{\lambda}) = e^{-1} - \bar{\lambda} \geq 0$ , and  $\psi(\rho, \bar{\lambda}) \rightarrow -h_0(\nu)(k_m)^{\nu+1} < 0$  as  $\rho \rightarrow 0$ . In addition, the function  $\psi$  is continuous and strictly monotonic in  $\rho$ . Thus, there is a unique solution  $\rho = \omega(\bar{\lambda}) \in (0, 1]$  for equation  $\psi(\rho, \bar{\lambda}) = 0$ .

c) For given  $\bar{\lambda} \in (0, e^{-1})$ ,  $\psi^{(0)}$  and  $\psi^{(2)}$  are both strictly increasing in  $\rho$ . In addition,  $\psi^{(0)} = \psi^{(2)} = 0$  when  $\rho = 1$ . Because the solution  $\rho = \omega(\bar{\lambda}) < 1$ , we have  $\psi^{(0)}(\omega(\bar{\lambda}), \bar{\lambda}) + \psi^{(2)}(\omega(\bar{\lambda}), \bar{\lambda}) < 0$  and thus  $\psi^{(1)}(\omega(\bar{\lambda}), \bar{\lambda}) = \omega(\bar{\lambda})e^{-\omega(\bar{\lambda})} - \bar{\lambda} > 0$ , i.e.,  $\omega(\bar{\lambda})e^{-\omega(\bar{\lambda})} > \bar{\lambda}$ .

#### APPENDIX E PROOF OF LEMMA 4

Since the  $T$ -slot drifts of  $X_t$  can be approximated by the drifts of  $X'_{t+s}$ , which can be further approximated by closed form expressions, we start our proof by examining the properties of these expressions and choose the values of  $\gamma$  and  $\delta$ . Then by choosing  $T$  and  $M$  properly, we make the approximate errors small enough such that the actual drifts have the same properties as their approximations.



Notice that  $\bar{\lambda} - \rho e^{-\rho} < 0$  when  $\rho = 1$  or  $\rho = \beta$ . In addition, it is a continuous function and is monotonically increasing in  $\rho \in [\beta, 1]$ . Therefore, there exist some  $\delta_1$  and  $\gamma$  such that  $\bar{\lambda} - \rho e^{-\rho} < \delta_1$  for all  $\rho \in [\beta - 5\gamma, 1 + 5\gamma]$ . For  $\Psi(\rho, \bar{\lambda})$ , with its strict monotonicity in  $\rho$ , we conclude that  $\psi(\rho, \bar{\lambda}) < \psi(\beta - \gamma, \bar{\lambda}) < 0$  for all  $\rho \in (0, \beta - \gamma)$  and  $\psi(\rho, \bar{\lambda}) > \psi(1 + \gamma, \bar{\lambda}) > 0$  for all  $\rho \in (1 + \gamma, \infty)$ . Thus, there exists some  $\delta$  such that

$$\bar{\lambda} - \rho e^{-\rho} \leq -3\delta/2, \quad \forall \rho \in [\beta - 5\gamma, 1 + 5\gamma], \quad (38)$$

$$\psi(\rho, \bar{\lambda}) \leq -3\delta, \quad \forall \rho \in (0, \beta - \gamma), \quad (39)$$

$$\psi(\rho, \bar{\lambda}) \geq 3\delta, \quad \forall \rho \in (1 + \gamma, \infty). \quad (40)$$

Now, fix  $\gamma$  and  $\delta$ . Using the uniform convergence of  $\frac{1}{T}\tilde{d}_T(n, \hat{n}, k)$ , we can choose a sufficiently large  $T$ , such that for any  $(n, \hat{n}, k) \in \mathbb{S}_X$ ,

$$\left| \frac{1}{T}\tilde{d}_T(n, \hat{n}, k) - \psi(\rho, \bar{\lambda}) \right| \leq \delta. \quad (41)$$

where  $\rho = n/\hat{n}$ .

According to Lemma 2, with the chosen  $T$ , there exists some  $M > 0$ , such that if  $n \geq M$  or  $\hat{n} \geq M$ , then

$$\left| \frac{1}{T}d_T(n, \hat{n}, k) - \frac{1}{T}d'_T(n, \hat{n}, k) \right| \leq \delta/2, \quad (42)$$

$$\left| \frac{1}{T}\hat{d}_T(n, \hat{n}, k) - \frac{1}{T}\hat{d}'_T(n, \hat{n}, k) \right| \leq \delta/2, \quad (43)$$

and thus,

$$\left| \frac{1}{T}\tilde{d}_T(n, \hat{n}, k) - \frac{1}{T}\tilde{d}'_T(n, \hat{n}, k) \right| \leq \delta. \quad (44)$$

With  $\frac{1}{T}d'_T(n, \hat{n}, k) = \bar{\lambda} - \rho e^{-\rho}$ , (38) and (42) together imply that for any  $(n, \hat{n}, k) \in S_{5\gamma, M}$ ,

$$\frac{1}{T}d_T(n, \hat{n}, k) \leq -\delta,$$

and thus (26) holds.

Similarly, (27) follows by combining (39), (41), and (44); (28) follows by combining (40), (41), and (44).

## REFERENCES

- [1] Cisco, "Cisco visual networking index: Global mobile data traffic forecast update, 2010-2015," Feb. 2011.
- [2] 3GPP TS 22.368 V11.0.2, "Service requirements for machine-type communications," Jun. 2011.
- [3] Huawei and CATR, "R2-100204: Traffic model for M2M services," in *3GPP TSG RAN WG2 Meeting #68bis*, Jan. 2010.
- [4] ZTE, "R2-104662: MTC simulation results with specific solutions," in *3GPP TSG RAN WG2 Meeting #71*, Aug. 2010.
- [5] P. Bertrand and J. Jiang, *LTE - The UMTS Long Term Evolution: From Theory to Practice*. John Wiley & Sons Ltd., 2009, ch. 19.
- [6] 3GPP TR 23.898 V7.0.0, "Access class barring and overload protection," Mar. 2005.
- [7] CATT, "R2-100182: Access control of MTC devices," in *3GPP TSG RAN WG2 Meeting #68bis*, Jan. 2010.
- [8] Telefon AB LM Ericsson, ST-Ericsson SA, Nokia Corporation, and Nokia, "GP-101378: Common assumptions for MTC simulation on CCCH and PDCH congestion," in *3GPP TSG GERAN #47*, 2010.
- [9] S.-Y. Lien, T.-H. Liao, C.-Y. Kao, and K.-C. Chen, "Cooperative access class barring for machine-to-machine communications," *IEEE Trans. Wireless Communications*, vol. 11, no. 1, pp. 27 – 32, Jan. 2012.
- [10] W. A. Rosenkrantz and D. Towsley, "On the instability of the slotted ALOHA multiaccess algorithm," *IEEE Trans. Automatic Control*, vol. 28, no. 10, pp. 994 – 996, Oct. 1983.
- [11] F. P. Kelly, "Stochastic models of communication systems," *Journal of the Royal Statistical Society (Series B)*, vol. 47, no. 3, pp. 379 – 395, 1985.
- [12] B. Hajek, "Hitting time and occupation time bounds implied by drift analysis with applications," *Advances in Applied Probability*, vol. 14, pp. 502 – 525, Sept. 1982.
- [13] J. N. Tsitsiklis, "Analysis of a multiaccess control scheme," *IEEE Trans. Automatic Control*, vol. 32, no. 11, pp. 1017 – 1020, Nov. 1987.
- [14] A. Kuczura, "The interrupted poisson process as an overflow process," *The Bell System Technical Journal*, vol. 52, no. 3, pp. 437 – 448, Mar 1973.
- [15] Telefon AB LM Ericsson and ST-Ericsson, "GP-101390: MTC device two stage access control," in *3GPP TSG GERAN #47*, Sept. 2010.
- [16] K.-R. Jung, A. Park, and S. Lee, "Machine-Type-Communication (MTC) device grouping algorithm for congestion avoidance of MTC oriented LTE network," *Communications in Computer and Information Science*, vol. 78, pp. 167 – 178, 2010.
- [17] R. Y. Kim, "Efficient wireless communications schemes for machine to machine communications," *Communications in Computer and Information Science*, vol. 181, no. 3, pp. 313 – 323, 2011.
- [18] S. D. Andreev, O. Galinina, and Y. Koucheryavy, "Envery-efficient client relay scheme for Machine-to-Machine communication," in *Proc. of IEEE GlobeCom 2011*, 2011.
- [19] G. Wang, X. Zhong, S. Mei, and J. Wang, "An adaptive medium access control mechanism for cellular based Machine to Machine (M2M) communication," in *Proc. of IEEE ICWITS 2010*, 2010.
- [20] G. A. Cunningham, III, "Delay versus throughput comparisons for stabilized slotted ALOHA," *IEEE Trans. Communications*, vol. 38, no. 11, pp. 1932 – 1934, Nov. 1990.
- [21] R. L. Rivest, "Network control by Bayesian broadcast," *IEEE Trans. Information Theory*, vol. 33, no. 3, pp. 323 – 328, May 1987.
- [22] B. Hajek and T. van Loon, "Decentralized dynamic control of a multiaccess broadcast channel," *IEEE Trans. Automatic Control*, vol. 27, no. 3, pp. 559 – 569, Jun. 1982.
- [23] ISO/IEC, "ISO/IEC 18000-6:2010 information technology - radio frequency identification for item management - part 6: Parameters for air interface communications at 860 MHz to 960 MHz," 2010.
- [24] D. Lee, K. Kim, and W. Lee, "Q<sup>+</sup>-algorithm: An enhanced RFID tag collision arbitration algorithm," *Lecture Notes in Computer Science, Ubiquitous Intelligence and Computing*, vol. 4611, no. 31, pp. 23 – 32, 2007.
- [25] H. Wu, C. Zhu, R. La, X. Liu, and Y. Zhang, "Fast adaptive S-ALOHA scheme for event-driven M2M communications," *IEEE VTC2012-Fall (Accepted)*. [Online]. Available: <http://arxiv.org/abs/1202.2998>
- [26] A. B. Carleial and M. E. Hellman, "Bistable behavior of ALOHA type systems," *IEEE Trans. Communications*, vol. 23, no. 4, pp. 401 – 410, Apr. 1975.
- [27] M. K. Gurcan and A. Al-Amir, "Joint drift analysis for multigroup slotted aloha: stability with maximum utilization," *IEEE Trans. Vehicular Technology*, vol. 50, no. 6, pp. 1415 – 1425, Nov. 2001.
- [28] M. Mitzenmacher and E. Upfal, *Probability and Computing: Randomized Algorithms and Probabilistic Analysis*. Cambridge University Press, 2005.